

Best Available Copy

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 September 2003 (12.09.2003)

(10) International Publication Number
WO 03/074740 A1

PCT

(51) International Patent Classification?: C12Q 1/68,
C07H 21/02, 21/04

CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE,
SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VC, VN, YU, ZA, ZM, ZW.

(21) International Application Number: PCT/US03/06376

(22) International Filing Date: 28 February 2003 (28.02.2003)

(25) Filing Language: English

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI,
SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN,
GQ, GW, ML, MR, NE, SN, TD, TG).

(26) Publication Language: English

(30) Priority Data:

| | | |
|------------|----------------------------|----|
| 60/360,232 | 1 March 2002 (01.03.2002) | US |
| 10/093,618 | 11 March 2002 (11.03.2002) | US |
| 60/378,354 | 8 May 2002 (08.05.2002) | US |

(71) Applicant and

(72) Inventor: DHALLAN, Ravinder [US/US]; 8013 Thorneley Court, Bethesda, MD 20817 (US).

(74) Agents: WITT, Eric et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 03/074740 A1

(54) Title: RAPID ANALYSIS OF VARIATIONS IN A GENOME

(57) Abstract: The invention provides a method useful for determining the sequence of large numbers of loci of interest on a single or multiple chromosomes. The method utilizes an oligonucleotide primer that contains a recognition site for a restriction enzyme such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest. The 5' overhang is used as a template to incorporate nucleotides, which can be detected. The method is especially amenable to the analysis of large numbers of sequences, such as single nucleotide polymorphisms, from one sample of nucleic acid.

RAPID ANALYSIS OF VARIATIONS IN A GENOME

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to U.S. Patent Application No. 10/093,618, filed March 11, 2002, and provisional U.S. Patent Application Nos. 60/360,232 and 60/378,354, filed March 1, 2002, and May 8, 2002, respectively. The contents of these applications are hereby incorporated by reference in their entirety herein.

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

[0002] The present invention is directed to a rapid method for determining the sequence of nucleic acid. The method is especially useful for genotyping, and for the detection of one to tens to hundreds to thousands of single nucleotide polymorphisms (SNPs) or mutations on single or on multiple chromosomes, and for the detection of chromosomal abnormalities, such as truncations, transversions, trisomies, and monosomies.

BACKGROUND

[0003] Sequence variation among individuals comprises a continuum from deleterious disease mutations to neutral polymorphisms. There are more than three thousand genetic diseases currently known including Duchenne Muscular Dystrophy, Alzheimer's Disease, Cystic Fibrosis, and Huntington's Disease (D.N. Cooper and M. Krawczak, "Human Genome Mutations," BIOS Scientific Publishers, Oxford (1993)). Also, particular DNA sequences may predispose individuals to a variety of diseases such as obesity, arteriosclerosis, and various types of cancer, including breast, prostate, and colon. In addition, chromosomal abnormalities, such as trisomy 21, which results in Down's Syndrome, trisomy 18, which results in Edward's Syndrome, trisomy 13, which results in Patau Syndrome, monosomy X, which results in Turner's Syndrome, and other sex aneuploidies, account for a significant portion of the genetic defects in liveborn

human beings. Knowledge of gene mutations, chromosomal abnormalities, and variations in gene sequences, such as single nucleotide polymorphisms (SNPs), will help to understand, diagnose, prevent, and treat diseases.

[0004] Most frequently, sequence variation is seen in differences in the lengths of repeated sequence elements, such as minisatellites and microsatellites, as small insertions or deletions, and as substitutions of the individual bases.

Single nucleotide polymorphisms (SNPs) represent the most common form of sequence variation; three million common SNPs with a population frequency of over 5% have been estimated to be present in the human genome. Small deletions or insertions, which usually cause frameshift mutations, occur on average, once in every 12 kilobases of genomic DNA (Wang, D.G. *et al.*, *Science* 280: 1077-1082 (1998)). A genetic map using these polymorphisms as a guide is being developed (<http://research.marshfieldclinic.org/genetics/>; internet address as of January 10, 2002).

[0005] The nucleic acid sequence of the human genome was published in February, 2001, and provides a genetic map of unprecedented resolution, containing several hundred thousand SNP markers, and a potential wealth of information on human diseases (Venter *et al.*, *Science* 291:1304-1351 (2001); International Human Genome Sequencing Consortium, *Nature* 409:860-921 (2001)). However, the length of DNA contained within the human chromosomes totals over 3 billion base pairs so sequencing the genome of every individual is impractical. Thus, it is imperative to develop high throughput methods for rapidly determining the presence of allelic variants of SNPs and point mutations, which predispose to or cause disease phenotypes. Efficient methods to characterize functional polymorphisms that affect an individual's physiology, psychology, audiology, ophthalmology, neurology, response to drugs, drug metabolism, and drug interactions also are needed.

[0006] Several techniques are widely used for analyzing and detecting genetic variations, such as DNA sequencing, restriction fragment length polymorphisms (RFLP), DNA hybridization assays, including DNA microarrays and peptide

nucleic acid analysis, and the Protein Truncation Test (PTT), all of which have limitations. Although DNA sequencing is the most definitive method, it is also the most time consuming and expensive. Often, the entire coding sequence of a gene is analyzed even though only a small fraction of the coding sequence is of interest. In most instances, a limited number of mutations in any particular gene account for the majority of the disease phenotypes.

[0007] For example, the cystic fibrosis transmembrane conductance regulator (CFTR) gene is composed of 24 exons spanning over 250,000 base pairs (Rommens et al., *Science* 245:1059-1065 (1989); Riordan et al., *Science* 245:1066-73 (1989)). Currently, there are approximately 200 mutations in the CFTR gene that are associated with a disease state of Cystic Fibrosis. Therefore, only a very small percentage of the reading frame for the CFTR gene needs to be analyzed. Furthermore, a total of 10 mutations make up 75.1% of all known disease cases. The deletion of a single phenylalanine residue, F508, accounts for 66% of all Cystic Fibrosis cases in Caucasians.

[0008] Hybridization techniques, including Southern Blots, Slot Blots, Dot Blots, and DNA microarrays, are commonly used to detect genetic variations (*Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Third Edition (2001)). In a typical hybridization assay, an unknown nucleotide sequence ("the target") is analyzed based on its affinity for another fragment with a known nucleotide sequence ("the probe"). If the two fragments hybridize under "stringent conditions," the sequences are thought to be complementary, and the sequence of the target fragment may be inferred from "the probe" sequence.

[0009] However, the results from a typical hybridization assay often are difficult to interpret. The absence or presence of a hybridization signal is dependent upon the definition of "stringent conditions." Any number of variables may be used to raise or lower stringency conditions such as salt concentration, the presence or absence of competitor nucleotide fragments, the number of washes performed to remove non-specific binding and the time and temperature at which the hybridizations are performed. Commonly, hybridization conditions must be

optimized for each "target" nucleotide fragment, which is time-consuming, and inconsistent with a high throughput method. A high degree of variability is often seen in hybridization assays, as well as a high proportion of false positives. Typically, hybridization assays function as a screen for likely candidates but a positive confirmation requires DNA sequencing analysis.

[0010] Several techniques for the detection of mutations have evolved based on the principal of hybridization analysis. For example, in the primer extension assay, the DNA region spanning the nucleotide of interest is amplified by PCR, or any other suitable amplification technique. After amplification, a primer is hybridized to a target nucleic acid sequence, wherein the last nucleotide of the 3' end of the primer anneals immediately 5' to the nucleotide position on the target sequence that is to be analyzed. The annealed primer is extended by a single, labeled nucleotide triphosphate. The incorporated nucleotide is then detected.

[0011] There are several limitations to the primer extension assay. First, the region of interest must be amplified prior to primer extension, which increases the time and expense of the assay. Second, PCR primers and dNTPs must be completely removed before primer extension, and residual contaminants can interfere with the proper analysis of the results. Third, and the most restrictive aspect of the assay, is that the primer is hybridized to the DNA template, which requires optimization of conditions for each primer, and for each sequence that is analyzed. Hybridization assays have a low degree of reproducibility, and a high degree of non-specificity.

[0012] The Peptide Nucleic Acid (PNA) affinity assay is a derivative of traditional hybridization assays (Nielsen et al., *Science* 254:1497-1500 (1991); Egholm et al., *J. Am. Chem. Soc.* 114:1895-1897 (1992); James et al., *Protein Science* 3:1347-1350 (1994)). PNAs are structural DNA mimics that follow Watson-Crick base pairing rules, and are used in standard DNA hybridization assays. PNAs display greater specificity in hybridization assays because a PNA/DNA mismatch is more destabilizing than a DNA/DNA mismatch and complementary PNA/DNA strands form stronger bonds than complementary

DNA/DNA strands. However, genetic analysis using PNAs still requires a laborious hybridization step, and as such, is subject to a high degree of non-specificity and difficulty with reproducibility.

[0013] Recently, DNA microarrays have been developed to detect genetic variations and polymorphisms (Taton et al., *Science* 289:1757-60, 2000; Lockhart et al., *Nature* 405:827-836 (2000); Gerhold et al., *Trends in Biochemical Sciences* 24:168-73 (1999); Wallace, R.W., *Molecular Medicine Today* 3:384-89 (1997); Blanchard and Hood, *Nature Biotechnology* 149:1649 (1996)). DNA microarrays are fabricated by high-speed robotics, on glass or nylon substrates, and contain DNA fragments with known identities ("the probe"). The microarrays are used for matching known and unknown DNA fragments ("the target") based on traditional base-pairing rules. The advantage of DNA microarrays is that one DNA chip may provide information on thousands of genes simultaneously. However, DNA microarrays are still based on the principle of hybridization, and as such, are subject to the disadvantages discussed above.

[0014] The Protein Truncation Test (PTT) is also commonly used to detect genetic polymorphisms (Roest et al., *Human Molecular Genetics* 2:1719-1721, (1993); Van Der Luit et al., *Genomics* 20:1-4 (1994); Hogervorst et al., *Nature Genetics* 10: 208-212 (1995)). Typically, in the PTT, the gene of interest is PCR amplified, subjected to in vitro transcription/translation, purified, and analyzed by polyacrylamide gel electrophoresis. The PTT is useful for screening large portions of coding sequence and detecting mutations that produce stop codons, which significantly diminish the size of the expected protein. However, the PTT is not designed to detect mutations that do not significantly alter the size of the protein.

[0015] Thus, a need still exists for a rapid method of analyzing DNA, especially genomic DNA suspected of having one or more single nucleotide polymorphisms or mutations.

BRIEF SUMMARY OF THE INVENTION

[0016] The invention is directed to a method for determining a sequence of a locus of interest, the method comprising: (a) amplifying a locus of interest on a template DNA using a first and second primers, wherein the second primer contains a recognition site for a restriction enzyme such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest; (b) digesting the amplified DNA with the restriction enzyme that recognizes the recognition site on the second primer; (c) incorporating a nucleotide into the digested DNA of (b) by using the 5' overhang containing the locus of interest as a template; and (d) determining the sequence of the locus of interest by determining the sequence of the DNA of (c).

[0017] The invention is also directed to a method for determining a sequence of a locus of interest, said method comprising: (a) amplifying a locus of interest on a template DNA using a first and second primers, wherein the second primer contains a portion of a recognition site for a restriction enzyme, wherein a full recognition site for the restriction enzyme is generated upon amplification of the template DNA such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest; (b) digesting the amplified DNA with the restriction enzyme that recognizes the full recognition site generated by the second primer and the template DNA; (c) incorporating a nucleotide into the digested DNA of (b) by using the 5' overhang containing the locus of interest as a template; and determining the sequence of the locus of interest by determining the sequence of the DNA of (c).

[0018] The invention also is directed to a method for determining a sequence of a locus of interest, said method comprising (a) replicating a region of DNA comprising a locus of interest from a template polynucleotide by using a first and a second primer, wherein the second primer contains a sequence that generates a recognition site for a restriction enzyme such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest; (b) digesting the DNA with the restriction enzyme that recognizes the recognition site generated by the second primer to create a DNA fragment; (c) incorporating a nucleotide into

the digested DNA of (b) by using the 5' overhang containing the locus of interest as a template; and (d) determining the sequence of the locus of interest by determining the sequence of the DNA of (c).

[0019] The invention also is directed to a DNA fragment containing a locus of interest to be sequenced and a recognition site for a restriction enzyme, wherein digestion with the restriction enzyme creates a 5' overhang on the DNA fragment, and wherein the locus of interest and the restriction enzyme recognition site are in relationship to each other such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest.

[0020] The template DNA can be obtained from any source including synthetic nucleic acid, preferably from a bacterium, fungus, virus, plant, protozoan, animal or human source. In one embodiment, the template DNA is obtained from a human source. In another embodiment, the template DNA is obtained from a cell, tissue, blood sample, serum sample, plasma sample, urine sample, spinal fluid, lymphatic fluid, semen, vaginal secretion, ascitic fluid, saliva, mucosa secretion, peritoneal fluid, fecal sample, or body exudates.

[0021] The 3' region of the first and/or second primer can contain a mismatch with the template DNA. The mismatch can occur at but is not limited to the last 1, 2, or 3 bases at the 3' end.

[0022] The restriction enzyme used in the invention can cut DNA at the recognition site. The restriction enzyme can be but is not limited to PflF I, Sau96 I, ScrF I, BsaJ I, Bssk I, Dde I, EcoN I, Fnu4H I, Hinf I, or Tth111 I. Alternatively, the restriction enzyme used in the invention can cut DNA at a distance from its recognition site.

[0023] In another embodiment, the first primer contains a recognition site for a restriction enzyme. In a preferred embodiment, the restriction enzyme recognition site is different from the restriction enzyme recognition site on the second primer. The invention includes digesting the amplified DNA with a restriction enzyme that recognizes the recognition site on the first primer.

[0024] Preferably, the recognition site on the second primer is for a restriction enzyme that cuts DNA at a distance from its recognition site and generates a 5' overhang, containing the locus of interest. In a preferred embodiment, the recognition site on the second primer is for a Type IIS restriction enzyme. The Type IIS restriction enzyme, e.g., is selected from the group consisting of: Alw I, Alw26 I, Bbs I, Bbv I, BceA I, Bmr I, Bsa I, Bst71 I, BsmA I, BsmB I, BsmF I, BspM I, Ear I, Fau I, Fok I, Hga I, Ple I, Sap I, SSfaN I, and Sthi32 I, and more preferably BceA I and BsmF I.

[0025] In one embodiment, the 5' region of the second primer does not anneal to the template DNA and/or the 5' region of the first primer does not anneal to the template DNA. The annealing length of the 3' region of the first or second primer can be 25-20, 20-15, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, or less than 4 bases.

[0026] In one embodiment, the amplification can comprise polymerase chain reaction (PCR). In a further embodiment, the annealing temperature for cycle 1 of PCR can be at about the melting temperature of the 3' region of the second primer that anneals to the template DNA. In another embodiment, the annealing temperature for cycle 2 of PCR can be about the melting temperature of the 3' region of the first primer that anneals to the template DNA. In another embodiment, the annealing temperature for the remaining cycles can be about the melting temperature of the entire sequence of the second primer.

[0027] In one embodiment, the 3' end of the second primer is adjacent to the locus of interest.

[0028] The first and/or second primer can contain a tag at the 5' terminus. Preferably, the first primer contains a tag at the 5' terminus. The tag can be used to separate the amplified DNA from the template DNA. The tag can be used to separate the amplified DNA containing the labeled nucleotide from the amplified DNA that does not contain the labeled nucleotide. The tag can be but is not limited to a radioisotope, fluorescent reporter molecule, chemiluminescent reporter molecule, antibody, antibody fragment, hapten, biotin, derivative of biotin, photobiotin, iminobiotin, digoxigenin, avidin, enzyme, acridinium, sugar,

enzyme, apoenzyme, homopolymeric oligonucleotide, hormone, ferromagnetic moiety, paramagnetic moiety, diamagnetic moiety, phosphorescent moiety, luminescent moiety, electrochemiluminescent moiety, chromatic moiety, moiety having a detectable electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity, or combinations thereof. Preferably, the tag is biotin. The biotin tag is used to separate amplified DNA from the template DNA using a streptavidin matrix. The streptavidin matrix is coated on wells of a microtiter plate.

[0029] The incorporation of a nucleotide in the method of the invention is by a DNA polymerase including but not limited to *E. coli* DNA polymerase, Klenow fragment of *E. coli* DNA polymerase I, T5 DNA polymerase, T7 DNA polymerase, T4 DNA polymerase, Taq polymerase, Pfu DNA polymerase, Vent DNA polymerase, bacteriophage 29, REDTaq™ Genomic DNA polymerase, and sequenase.

[0030] The incorporation of a nucleotide can further comprise using a mixture of labeled and unlabeled nucleotides. One nucleotide, two nucleotides, three nucleotides, four nucleotides, five nucleotides, or more than five nucleotides may be incorporated. A combination of labeled and unlabeled nucleotides can be incorporated. The labeled nucleotide can be but is not limited to a dideoxynucleotide triphosphate and deoxynucleotide triphosphate. The unlabeled nucleotide can be but is not limited to a dideoxynucleotide triphosphate and deoxynucleotide triphosphate. The labeled nucleotide is labeled with a molecule such as but not limited to a radioactive molecule, fluorescent molecule, antibody, antibody fragment, hapten, carbohydrate, biotin, and derivative of biotin, phosphorescent moiety, luminescent moiety, electrochemiluminescent moiety, chromatic moiety, or moiety having a detectable electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity. Preferably, the labeled nucleotide is labeled with a fluorescent molecule. The incorporation of a fluorescent labeled nucleotide further includes using a mixture of fluorescent and unlabeled nucleotides.

[0031] In one embodiment, the determination of the sequence of the locus of interest comprises detecting the incorporated nucleotide. In one embodiment, the detection is by a method such as but not limited to gel electrophoresis, capillary electrophoresis, microchannel electrophoresis, polyacrylamide gel electrophoresis, fluorescence detection, sequencing, ELISA, mass spectrometry, time of flight mass spectrometry, quadrupole mass spectrometry, magnetic sector mass spectrometry, electric sector mass spectrometry, fluorometry, infrared spectrometry, ultraviolet spectrometry, palentiostatic amperometry, hybridization, such as Southern Blot, or microarray. In a preferred embodiment, the detection is by fluorescence detection.

[0032] In a preferred embodiment, the locus of interest is suspected of containing a single nucleotide polymorphism or mutation. The method can be used for determining sequences of multiple loci of interest concurrently. The template DNA can comprise multiple loci from a single chromosome. The template DNA can comprise multiple loci from different chromosomes. The loci of interest on template DNA can be amplified in one reaction. Alternatively, each of the loci of interest on template DNA can be amplified in a separate reaction. The amplified DNA can be pooled together prior to digestion of the amplified DNA. Each of the labeled DNA containing a locus of interest can be separated prior to determining the sequence of the locus of interest. In one embodiment, at least one of the loci of interest is suspected of containing a single nucleotide polymorphism or a mutation.

[0033] In another embodiment, the method of the invention can be used for determining the sequences of multiple loci of interest from a single individual or from multiple individuals. Also, the method of the invention can be used to determine the sequence of a single locus of interest from multiple individuals.

BRIEF DESCRIPTION OF THE FIGURES

[0034] FIG. 1A. A Schematic diagram depicting a double stranded DNA molecule. A pair of primers, depicted as bent arrows, flank the locus of interest, depicted as a triangle symbol at base N14. The locus of interest can be a single

nucleotide polymorphism, point mutation, insertion, deletion, translocation, etc. Each primer contains a restriction enzyme recognition site about 10 bp from the 5' terminus depicted as region "a" in the first primer and as region "d" in the second primer. Restriction recognition site "a" can be for any type of restriction enzyme but recognition site "d" is for a restriction enzyme, which cuts "n" nucleotides away from its recognition site and leaves a 5' overhang and a recessed 3' end. Examples of such enzymes include but are not limited to BceA I and BsmF I. The 5' overhang serves as a template for incorporation of a nucleotide into the 3' recessed end.

[0035] The first primer is shown modified with biotin at the 5' end to aid in purification. The sequence of the 3' end of the primers is such that the primers anneal at a desired distance upstream and downstream of the locus of interest. The second primer anneals close to the locus of interest; the annealing site, which is depicted as region "c," is designed such that the 3' end of the second primer anneals one base away from the locus of interest. The second primer can anneal any distance from the locus of interest provided that digestion with the restriction enzyme, which recognizes the region "d" on this primer, generates a 5' overhang that contains the locus of interest.

[0036] The first primer annealing site, which is depicted as region "b," is about 20 bases.

[0037] FIG. 1B. A schematic diagram depicting the annealing and extension steps of the first cycle of amplification by PCR. The first cycle of amplification is performed at about the melting temperature of the 3' region, which anneals to the template DNA, of the second primer, depicted as region "c," and is 13 base pairs in this example. At this temperature, both the first and second primers anneal to their respective complementary strands and begin extension, depicted by dotted lines. In this first cycle, the second primer extends and copies the region b where the first primer can anneal in the next cycle.

[0038] FIG. 1C. A schematic diagram depicting the annealing and extension steps following denaturation in the second cycle of amplification of PCR. The

second cycle of amplification is performed at a higher annealing temperature (TM2), which is about the melting temperature of the 20 bp of the 3' region of the first primer that anneals to the template DNA, depicted as region "b." Therefore at TM2, the first primer, which is complementary to region b, can bind to the DNA that was copied in the first cycle of the reaction. However, at TM2 the second primer cannot anneal to the original template DNA or to DNA that was copied in the first cycle of the reaction because the annealing temperature is too high. The second primer can anneal to 13 bases in the original template DNA but TM2 is calculated at about the melting temperature of 20 bases.

[0039] FIG. 1D. A schematic diagram depicting the annealing and extension reactions after denaturation during the third cycle of amplification. In this cycle, the annealing temperature, TM3, is about the melting temperature of the entire second primer, including regions "c" and "d." The length of regions "c" + "d" is about 27-33 bp long, and thus TM3 is significantly higher than TM1 and TM2. At this higher TM the second primer, which contain region c and d, anneals to the copied DNA generated in cycle 2.

[0040] FIG. 1E. A schematic diagram depicting the annealing and extension reactions for the remaining cycles of amplification. The annealing temperature for the remaining cycles is TM3, which is about the melting temperature of the entire second primer. At TM3, the second primer binds to templates that contain regions c' and d' and the first primer binds to templates that contain regions a' and b. By raising the annealing temperature successively in each cycle for the first three cycles, from TM1 to TM2 to TM3, nonspecific amplification is significantly reduced.

[0041] FIG. 1F. A schematic diagram depicting the amplified locus of interest bound to a solid matrix.

[0042] FIG. 1G. A schematic diagram depicting the bound, amplified DNA after digestion with a restriction enzyme that recognizes "d." The "downstream" end is released into the supernatant, and can be removed by washing with any

suitable buffer. The upstream end containing the locus of interest remains bound to the solid matrix.

[0043] FIG. 1H. A schematic diagram depicting the bound amplified DNA, after "filling in" with a labeled ddNTP. A DNA polymerase is used to "fill in" the base (N'_{14}) that is complementary to the locus of interest (N_{14}). In this example, only ddNTPs are present in this reaction, such that only the locus of interest or SNP of interest is filled in.

[0044] FIG. 1I. A schematic diagram depicting the labeled, bound DNA after digestion with restriction enzyme "a." The labeled DNA is released into the supernatant, which can be collected to identify the base that was incorporated.

[0045] FIG. 2. A schematic diagram depicting double stranded DNA templates with "N" number of loci of interest and "n" number of primer pairs, x_1, y_1 to x_n, y_n , specifically annealed such that a primer flanks each locus of interest. The first primers are biotinylated at the 5' end, depicted by •, and contain a restriction enzyme recognition site, "a", which is recognized by any type of restriction enzyme. The second primers contain a restriction enzyme recognition site, "d," where "d" is a recognition site for a restriction enzyme that cuts DNA at a distance from its recognition site, and generates a 5' overhang containing the locus of interest and a recessed 3' end. The second primers anneal adjacent to the respective loci of interest. The exact position of the restriction enzyme site "d" in the second primers is designed such that digesting the PCR product of each locus of interest with restriction enzyme "d" generates a 5' overhang containing the locus of interest and a 3' recessed end. The annealing sites of the first primers are about 20 bases long and are selected such that each successive first primer is further away from its respective second primer. For example, if at locus 1 the 3' ends of the first and second primers are Z base pairs apart, then at locus 2, the 3' ends of the first and second primers are $Z + K$ base pairs apart, where $K = 1, 2, 3$ or more than three bases. Primers for locus N are $Z_{N-1} + K$ base pairs apart. The purpose of making each successive first primer further apart from their respective second primers is such that the "filled in" restriction fragments (generated after

amplification, purification, digestion and labeling as described in FIGS. 1B-1I) differ in size and can be resolved, for example by electrophoresis, to allow detection of each individual locus of interest.

[0046] FIG. 3. PCR amplification of DNA fragments containing SNPs using multiple annealing temperatures. A sample containing genomic DNA templates from thirty-six human volunteers was analyzed for the following four SNPs: SNP HC21S00340 (lane 1), identification number as assigned in the Human Chromosome 21 cSNP Database, located on chromosome 21; SNP TSC 0095512 (lane 2), located on chromosome 1; SNP TSC 0214366 (lane 3), located on chromosome 1; and SNP TSC 0087315 (lane 4), located on chromosome 1. Each DNA fragment containing a SNP was amplified by PCR using three different annealing temperature protocols, herein referred to as the low stringency annealing temperature; medium stringency annealing temperature; and high stringency annealing temperature. Regardless of the annealing temperature protocol, each DNA fragment containing a SNP was amplified for 40 cycles of PCR. The denaturation step for each PCR reaction was performed for 30 seconds at 95°C.

[0047] FIG. 3A. Photograph of a gel demonstrating PCR amplification of the 4 DNA fragments containing different SNPs using the low stringency annealing temperature protocol.

[0048] FIG. 3B. Photograph of a gel demonstrating PCR amplification of the 4 DNA fragments containing different SNPs using the medium stringency annealing temperature protocol.

[0049] FIG. 3C. Photograph of a gel demonstrating PCR amplification of the 4 DNA fragments containing different SNPs using the high stringency annealing temperature protocol.

[0050] FIG. 4A. A depiction of the DNA sequence of SNP HC21S00027 (SEQ ID NOS:27 & 28), assigned by the Human Chromosome 21 cSNP database, located on chromosome 21. A first primer (SEQ ID NO:17) and a second primer (SEQ ID NO:18) are indicated above and below, respectively, the

sequence of HC21S00027. The first primer is biotinylated and contains the restriction enzyme recognition site for EcoRI. The second primer contains the restriction enzyme recognition site for BsmF I and contains 13 bases that anneal to the DNA sequence. The SNP is indicated by R (A/G) and r (T/C; complementary to R).

[0051] FIG. 4B. A depiction of the DNA sequence of SNP HC21S00027 (SEQ ID NOS:27 & 28), as assigned by the Human Chromosome 21 cSNP database, located on chromosome 21. A first primer (SEQ ID NO:17) and a second primer (SEQ ID NO:19) are indicated above and below, respectively, the sequence of HC21S00027. The first primer is biotinylated and contains the restriction enzyme recognition site for EcoRI. The second primer contains the restriction enzyme recognition site for BceA I and has 13 bases that anneal to the DNA sequence. The SNP is indicated by R (A/G) and r (T/C; complementary to R).

[0052] FIG. 4C. A depiction of the DNA sequence of SNP TSC0095512 (SEQ ID NOS:29 & 30) from chromosome 1. The first primer (SEQ ID NO:11) and the second primer (SEQ ID NO:20) are indicated above and below, respectively, the sequence of TSC0095512. The first primer is biotinylated and contains the restriction enzyme recognition site for EcoRI. The second primer contains the restriction enzyme recognition site for BsmF I and has 13 bases that anneal to the DNA sequence. The SNP is indicated by S (G/C) and s (C/G; complementary to S).

[0053] FIG. 4D. A depiction of the DNA sequence of SNP TSC0095512 (SEQ ID NOS:29 & 30) from chromosome 1. The first primer (SEQ ID NO:11) and the second primer (SEQ ID NO:12) are indicated above and below, respectively, the sequence of TSC0095512. The first primer is biotinylated and contains the restriction enzyme recognition site for EcoRI. The second primer contains the restriction enzyme recognition site for BceA I and has 13 bases that anneal to the DNA sequence. The SNP is indicated by S (G/C) and s (C/G; complementary to S).

[0054] FIGS. 5A-5D. A schematic diagram depicting the nucleotide sequences of SNP HC21S00027 (FIG. 5A (SEQ ID NOS:31 & 32) and FIG. 5B (SEQ ID NOS:31 & 33)), and SNP TSC0095512 (FIG. 5C (SEQ ID NOS:34 & 35) and FIG. 5D (SEQ ID NOS:34 & 36)) after amplification with the primers described in FIGS. 4A-4D. Restriction sites in the primer sequence are indicated in bold.

[0055] FIGS. 6A-6D. A schematic diagram depicting the nucleotide sequences of each amplified DNA fragment containing a SNP after digestion with the appropriate Type IIS restriction enzyme. FIG. 6A (SEQ ID NOS:31 & 32) and FIG. 6B (SEQ ID NOS:31 & 33) depict fragments of a DNA sequence containing SNP HC21S00027 digested with the Type IIS restriction enzymes BsmF I and BceA I, respectively. FIG. 6C (SEQ ID NOS:34 & 35) and FIG. 6D (SEQ ID NOS:34 & 36) depict fragments of a DNA sequence containing SNP TSC0095512 digested with the Type IIS restriction enzymes BsmF I and BceA I, respectively.

[0056] FIGS. 7A-7D. A schematic diagram depicting the incorporation of a fluorescently labeled nucleotide using the 5' overhang of the digested SNP site as a template to "fill in" the 3' recessed end. FIG. 7A (SEQ ID NOS:31, 37 & 41) and FIG. 7B (SEQ ID NOS:31, 37 & 39) depict the digested SNP HC21S00027 locus with an incorporated labeled ddNTP (*R^{-dd} = fluorescent dideoxy nucleotide). FIG. 7C (SEQ ID NOS:34 & 38) and FIG. 7D (SEQ ID NO:34) depict the digested SNP TSC0095512 locus with an incorporated labeled ddNTP (*S^{-dd} = fluorescent dideoxy nucleotide). The use of ddNTPs ensures that the 3' recessed end is extended by one nucleotide, which is complementary to the nucleotide of interest or SNP site present in the 5' overhang.

[0057] FIG. 7E. A schematic diagram depicting the incorporation of dNTPs and a ddNTP into the 5' overhang containing the SNP site. The DNA fragment containing SNP HC21S00007 was digested with BsmF I, which generates a four base 5' overhang. The use of a mixture of dNTPs and ddNTPs allows the 3' recessed end to be extended one nucleotide (a ddNTP is incorporated first) (SEQ

ID NOS:31, 37 & 41); two nucleotides (a dNTP is incorporated followed by a ddNTP) (SEQ ID NOS:31, 39 & 41); three nucleotides (two dNTPs are incorporated, followed by a ddNTP) (SEQ ID NOS:31, 40 & 41); or four nucleotides (three dNTPs are incorporated, followed by a ddNTP) (SEQ ID NOS:31 & 41). All four products can be separated by size, and the incorporated nucleotide detected (*R^{dd} = fluorescent dideoxy nucleotide). Detection of the first nucleotide, which corresponds to the SNP or locus site, and the next three nucleotides provides an additional level of quality assurance. The SNP is indicated by R (A/G) and r (T/C) (complementary to R).

[0058] FIGS. 8A-8D. Release of the "filled in" SNP from the solid support matrix, i.e. streptavidin coated well. SNP HC21S00027 is shown in FIG. 8A (SEQ ID NOS:31, 37 & 41) and FIG. 8B (SEQ ID NOS:31, 37 & 39), while SNP TSC0095512 is shown in FIG. 8C (SEQ ID NOS:34 & 38) and FIG. 8D (SEQ ID NO:34). The "filled in" SNP is free in solution, and can be detected.

[0059] FIG. 9A. Sequence analysis of a DNA fragment containing SNP HC21S00027 digested with BceAI. Four "fill in" reactions are shown; each reaction contained one fluorescently labeled nucleotide, ddGTP, ddATP, ddTTP, or ddCTP, and unlabeled ddNTPs. The 5' overhang generated by digestion with BceA I and the expected nucleotides at this SNP site are indicated.

[0060] FIG 9B. Sequence analysis of SNP TSC0095512. SNP TSC0095512 was amplified with a second primer that contained the recognition site for BceA I, and in a separate reaction, with a second primer that contained the recognition site for BsmF I. Four fill in reactions are shown for each PCR product; each reaction contained one fluorescently labeled nucleotide, ddGTP, ddATP, ddTTP, or ddCTP, and unlabeled ddNTPs. The 5' overhang generated by digestion with BceA I and with BsmF I and the expected nucleotides are indicated.

[0061] FIG 9C. Sequence analysis of SNP TSC0264580 after amplification with a second primer that contained the recognition site for BsmF I. Four "fill in" reactions are shown; each reaction contained one fluorescently labeled nucleotide, which was ddGTP, ddATP, ddTTP, or ddCTP and unlabeled ddNTPs. Two

different 5' overhangs are depicted: one represents the DNA molecules that were cut 11 nucleotides away on the sense strand and 15 nucleotides away on the antisense strand and the other represents the DNA molecules that were cut 10 nucleotides away on the sense strand and 14 nucleotides away on the antisense strand. The expected nucleotides also are indicated.

[0062] FIG 9D. Sequence analysis of SNP HC21S00027 amplified with a second primer that contained the recognition site for BsmF I. A mixture of labeled ddNTPs and unlabeled dNTPs was used to fill in the 5' overhang generated by digestion with BsmF I. Two different 5' overhangs are depicted: one represents the DNA molecules that were cut 11 nucleotides away on the sense strand and 15 nucleotides away on the antisense strand and the other represents the DNA molecules that were cut 10 nucleotides away on the sense strand and 14 nucleotides away on the antisense strand. The nucleotide upstream of the SNP, the nucleotide at the SNP site (the sample contained DNA templates from 36 individuals; both nucleotides would be expected to be represented in the sample), and the three nucleotides downstream of the SNP are indicated.

[0063] FIG. 10. Sequence analysis of multiple SNPs. SNPs HC21S00131, and HC21S00027, which are located on chromosome 21, and SNPs TSC0087315, SNP TSC0214366, SNP TSC0413944, and SNP TSC0095512, which are on chromosome 1, were amplified in separate PCR reactions with second primers that contained a recognition site for BsmF I. The primers were designed so that each amplified locus of interest was of a different size. After amplification, the reactions were pooled into a single sample, and all subsequent steps of the method performed (as described for FIGS. 1F-1I) on that sample. Each SNP and the nucleotide found at each SNP are indicated.

[0064] FIG. 11. Sequence determination of both alleles of SNPs TSC0837969, TSC0034767, TSC1130902, TSC0597888, TSC0195492, TSC0607185 using one fluorescently labeled nucleotide. Labeled ddGTP was used in the presence of unlabeled dATP, dCTP, dTTP to fill-in the overhang generated by digestion with BsmF I. The nucleotide preceding the variable site on

the strand that was filled-in was not guanine, and the nucleotide after the variable site on the strand that was filled in was not guanine. The nucleotide two bases after the variable site on the strand that was filled-in was guanine. Alleles that contain guanine at variable site are filled in with labeled ddGTP. Alleles that do not contain guanine are filled in with unlabeled dATP, dCTP, or dTTP, and the polymerase continues to incorporate nucleotides until labeled ddGTP is filled in at position 3 complementary to the overhang.

DETAILED DESCRIPTION OF THE INVENTION

[0065] The present invention provides a novel method for rapidly determining the sequence of DNA, especially at a locus of interest or multiple loci of interest. The sequences of any number of DNA targets, from one to hundreds or thousands or more of loci of interest in any template DNA or sample of nucleic acid can be determined efficiently, accurately, and economically. The method is especially useful for the rapid sequencing of one to tens of thousands or more of genes, regions of genes, fragments of genes, single nucleotide polymorphisms, and mutations on a single chromosome or on multiple chromosomes.

[0066] The invention is directed to a method for determining a sequence of a locus of interest, the method comprising: (a) amplifying a locus of interest on a template DNA using a first and second primers, wherein the second primer contains a recognition site for a restriction enzyme such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest; (b) digesting the amplified DNA with the restriction enzyme that recognizes the recognition site on the second primer; (c) incorporating a nucleotide into the digested DNA of (b) by using the 5' overhang containing the locus of interest as a template; and (d) determining the sequence of the locus of interest by determining the sequence of the DNA of (c).

[0067] The invention is also directed to a method for determining a sequence of a locus of interest, said method comprising: (a) amplifying a locus of interest on a template DNA using a first and second primers, wherein the first and/or second primer contains a portion of a recognition site for a restriction enzyme,

wherein a full recognition site for the restriction enzyme is generated upon amplification of the template DNA such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest; (b) digesting the amplified DNA with the restriction enzyme that recognizes the full recognition site generated by the second primer and the template DNA; (c) incorporating a nucleotide into the digested DNA of (b) by using the 5' overhang containing the locus of interest as a template; and determining the sequence of the locus of interest by determining the sequence of the DNA of (c).

DNA Template

[0068] By a "locus of interest" is intended a selected region of nucleic acid that is within a larger region of nucleic acid. A locus of interest can include but is not limited to 1-100, 1-50, 1-20, or 1-10 nucleotides, preferably 1-6, 1-5, 1-4, 1-3, 1-2, or 1 nucleotide(s).

[0069] As used herein, an "allele" is one of several alternate forms of a gene or non-coding regions of DNA that occupy the same position on a chromosome. The term allele can be used to describe DNA from any organism including but not limited to bacteria, viruses, fungi, protozoa, molds, yeasts, plants, humans, non-humans, animals, and archaebacteria.

[0070] As used herein with respect to individuals, "mutant alleles" refers to variant alleles that are associated with a disease state.

For example, bacteria typically have one large strand of DNA. The term allele with respect to bacterial DNA refers to the form of a gene found in one cell as compared to the form of the same gene in a different bacterial cell of the same species.

[0071] Alleles can have the identical sequence or can vary by a single nucleotide or more than one nucleotide. With regard to organisms that have two copies of each chromosome, if both chromosomes have the same allele, the condition is referred to as homozygous. If the alleles at the two chromosomes are different, the condition is referred to as heterozygous. For example, if the locus of interest is SNP X on chromosome 1, and the maternal chromosome contains an

adenine at SNP X (A allele) and the paternal chromosome contains a guanine at SNP X (G allele), the individual is heterozygous at SNP X.

[0072] As used herein, "sequence" means the identity of, or to determine the identity of (depending on whether used as a noun or a verb, respectively), one nucleotide or more than one contiguous nucleotides in a polynucleotide. In the case of a single nucleotide, e.g., a SNP, "sequence" is used as a noun interchangeably with "identity" herein, and "sequence" is used interchangeably as a verb with "identify" herein.

[0073] The term "template" refers to any nucleic acid molecule that can be used for amplification in the invention. RNA or DNA that is not naturally double stranded can be made into double stranded DNA so as to be used as template DNA. Any double stranded DNA or preparation containing multiple, different double stranded DNA molecules can be used as template DNA to amplify a locus or loci of interest contained in the template DNA.

[0074] The source of the nucleic acid for obtaining the template DNA can be from any appropriate source including but not limited to nucleic acid from any organism, e.g., human or nonhuman, e.g., bacterium, virus, yeast, fungus, plant, protozoan, animal, nucleic acid-containing samples of tissues, bodily fluids (for example, blood, serum, plasma, saliva, urine, tears, semen, vaginal secretions, lymph fluid, cerebrospinal fluid or mucosa secretions), fecal matter, individual cells or extracts of the such sources that contain the nucleic acid of the same, and subcellular structures such as mitochondria or chloroplasts, using protocols well established within the art. Nucleic acid can

[same paragraph] also be obtained from forensic, food, archeological, or inorganic samples onto which nucleic acid has been deposited or extracted. In a preferred embodiment, the nucleic acid has been obtained from a human or animal to be screened for the presence of one or more genetic sequences that can be diagnostic for, or predispose the subject to, a medical condition or disease.

[0075] The nucleic acid that is to be analyzed can be any nucleic acid, e.g., genomic, plasmid, cosmid, yeast artificial chromosomes, artificial or man-made

DNA, including unique DNA sequences, and also DNA that has been reverse transcribed from an RNA sample, such as cDNA. The sequence of RNA can be determined according to the invention if it is capable of being made into a double stranded DNA form to be used as template DNA.

[0076] The terms "primer" and "oligonucleotide primer" are interchangeable when used to discuss an oligonucleotide that anneals to a template and can be used to prime the synthesis of a copy of that template.

[0077] "Amplified" DNA is DNA that has been "copied" once or multiple times, e.g. by polymerase chain reaction. When a large amount of DNA is available to assay, such that a sufficient number of copies of the locus of interest are already present in the sample to be assayed, it may not be necessary to "amplify" the DNA of the locus of interest into an even larger number of replicate copies. Rather, simply "copying" the template DNA once using a set of appropriate primers, such as those containing hairpin structures that allow the restriction enzyme recognition sites to be double stranded, can suffice.

[0078] "Copy" as in "copied DNA" refers to DNA that has been copied once, or DNA that has been amplified into more than one copy.

[0079] In one embodiment, the nucleic acid is amplified directly in the original sample containing the source of nucleic acid. It is not essential that the nucleic acid be extracted, purified or isolated; it only needs to be provided in a form that is capable of being amplified. A hybridization step of the nucleic acid with the primers, prior to amplification, is not required. For example, amplification can be performed in a cell or sample lysate using standard protocols well known in the art. DNA that is on a solid support, in a fixed biological preparation, or otherwise in a composition that contains non-DNA substances and that can be amplified without first being extracted from the solid support or fixed preparation or non-DNA substances in the composition can be used directly, without further purification, as long as the DNA can anneal with appropriate primers, and be copied, especially amplified, and the copied or amplified products can be recovered and utilized as described herein.

[0080] In a preferred embodiment, the nucleic acid is extracted, purified or isolated from non-nucleic acid materials that are in the original sample using methods known in the art prior to amplification.

[0081] In another embodiment, the nucleic acid is extracted, purified or isolated from the original sample containing the source of nucleic acid and prior to amplification, the nucleic acid is fragmented using any number of methods well known in the art including but not limited to enzymatic digestion, manual shearing, and sonication. For example, the DNA can be digested with one or more restriction enzymes that have a recognition site, and especially an eight base or six base pair recognition site, which is not present in the loci of interest. Typically, DNA can be fragmented to any desired length, including 50, 100, 250, 500, 1,000, 5,000, 10,000, 50,000 and 100,000 base pairs long. In another embodiment, the DNA is fragmented to an average length of about 1000 to 2000 base pairs. However, it is not necessary that the DNA be fragmented.

[0082] Fragments of DNA that contain the loci of interest can be purified from the fragments of DNA that do not contain the loci of interest before amplification. The purification can be done by using primers that will be used in the amplification (see "Primer Design" section below) as hooks to retrieve the fragments containing the loci of interest, based on the ability of such primers to anneal to the loci of interest. In a preferred embodiment, tag-modified primers are used, such as e.g. biotinylated primers. See also the "Purification of Amplified DNA" section for additional tags.

[0083] By purifying the DNA fragments containing the loci of interest, the specificity of the amplification reaction can be improved. This will minimize amplification of nonspecific regions of the template DNA. Purification of the DNA fragments can also allow multiplex PCR (Polymerase Chain Reaction) or amplification of multiple loci of interest with improved specificity.

[0084] In one embodiment, the nucleic acid sample is obtained with a desired purpose in mind such as to determine the sequence at a predetermined locus or loci of interest using the method of the invention. For example, the nucleic acid is

obtained for the purpose of identifying one or more conditions or diseases to which the subject can be predisposed or is in need of treatment for, or the presence of certain single nucleotide polymorphisms. In an alternative embodiment, the sample is obtained to screen for the presence or absence of one or more DNA sequence markers, the presence of which would identify that DNA as being from a specific bacterial or fungal microorganism, or individual.

[0085] The loci of interest that are to be sequenced can be selected based upon sequence alone. In humans, over 1.42 million single nucleotide polymorphisms (SNPs) have been described (Nature 409:928-933 (2001); The SNP Consortium LTD). On the average, there is one SNP every 1.9 kb of human genome. However, the distance between loci of interest need not be considered when selecting the loci of interest to be sequenced according to the invention. If more than one locus of interest on genomic DNA is being analyzed, the selected loci of interest can be on the same chromosome or on different chromosomes.

[0086] In a preferred embodiment, the length of sequence that is amplified is preferably different for each locus of interest so that the loci of interest can be separated by size.

[0087] In fact, it is an advantage of the invention that primers that copy an entire gene sequence need not be utilized. Rather, the copied locus of interest is preferably only a small part of the total gene. There is no advantage to sequencing the entire gene as this can increase cost and delay results. Sequencing only the desired bases or loci of interest within the gene maximizes the overall efficiency of the method because it allows for the maximum number of loci of interest to be determined in the fastest amount of time and with minimal cost.

[0088] Because a large number of sequences can be analyzed together, the method of the invention is especially amenable to the large-scale screening of a number of individual samples.

[0089] Any number of loci of interest can be analyzed and processed, especially concurrently, using the method of the invention. The sample(s) can be analyzed to determine the sequence at one locus of interest or at multiple loci of

interest concurrently. For example, the 10 or 20 most frequently occurring mutation sites in a disease associated gene can be sequenced to detect the majority of the disease carriers.

[0090] Alternatively, 2, 3, 4, 5, 6, 7, 8, 9, 10-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,000, 2,000-3,000, 3,000-5,000, 5,000-10,000, 10,000-50,000 or more than 50,000 loci of interest can be analyzed at the same time when a global genetic screening is desired. Such a global genetic screening might be desired when using the method of the invention to provide a genetic fingerprint to identify a certain microorganism or individual or for SNP genotyping.

[0091] The multiple loci of interest can be targets from different organisms. For example, a plant, animal or human subject in need of treatment can have symptoms of infection by one or more pathogens. A nucleic acid sample taken from such a plant, animal or human subject can be analyzed for the presence of multiple suspected or possible pathogens at the same time by determining the sequence of loci of interest which, if present, would be diagnostic for that pathogen. Not only would the finding of such a diagnostic sequence in the subject rapidly pinpoint the cause of the condition, but also it would rule out other pathogens that were not detected. Such screening can be used to assess the degree to which a pathogen has spread throughout an organism or environment. In a similar manner, nucleic acid from an individual suspected of having a disease that is the result of a genetic abnormality can be analyzed for some or all of the known mutations that result in the disease, or one or more of the more common mutations.

[0092] The method of the invention can be used to monitor the integrity of the genetic nature of an organism. For example, samples of yeast can be taken at various times and from various batches in the brewing process, and their presence or identity compared to that of a desired strain by the rapid analysis of their genomic sequences as provided herein.

[0093] The locus of interest that is to be copied can be within a coding sequence or outside of a coding sequence. Preferably, one or more loci of interest that are to be copied are within a gene. In a preferred embodiment, the template DNA that is copied is a locus or loci of interest that is within a genomic coding sequence, either intron or exon. In a highly preferred embodiment, exon DNA sequences are copied. The loci of interest can be sites where mutations are known to cause disease or predispose to a disease state. The loci of interest can be sites of single nucleotide polymorphisms. Alternatively, the loci of interest that are to be copied can be outside of the coding sequence, for example, in a transcriptional regulatory region, and especially a promoter, enhancer, or repressor sequence.

Primer Design

[0094] Published sequences, including consensus sequences, can be used to design or select primers for use in amplification of template DNA. The selection of sequences to be used for the construction of primers that flank a locus of interest can be made by examination of the sequence of the loci of interest, or immediately thereto. The recently published sequence of the human genome provides a source of useful consensus sequence information from which to design primers to flank a desired human gene locus of interest.

[0095] By "flanking" a locus of interest is meant that the sequences of the primers are such that at least a portion of the 3' region of one primer is complementary to the antisense strand of the template DNA and upstream of the locus of interest (forward primer), and at least a portion of the 3' region of the other primer is complementary to the sense strand of the template DNA and downstream of the locus of interest (reverse primer). A "primer pair" is intended to specify a pair of forward and reverse primers. Both primers of a primer pair anneal in a manner that allows extension of the primers, such that the extension results in amplifying the template DNA in the region of the locus of interest.

[0096] Primers can be prepared by a variety of methods including but not limited to cloning of appropriate sequences and direct chemical synthesis using methods well known in the art (Narang et al., Methods Enzymol. 68:90 (1979);

Brown et al., *Methods Enzymol.* 68:109 (1979)). Primers can also be obtained from commercial sources such as Operon Technologies, Amersham Pharmacia Biotech, Sigma, and Life Technologies. The primers of a primer pair can have the same length. Alternatively, one of the primers of the primer pair can be longer than the other primer of the primer pair. The primers can have an identical melting temperature. The lengths of the primers can be extended or shortened at the 5' end or the 3' end to produce primers with desired melting temperatures. In a preferred embodiment, the 3' annealing lengths of the primers, within a primer pair, differ. Also, the annealing position of each primer pair can be designed such that the sequence and length of the primer pairs yield the desired melting temperature. The simplest equation for determining the melting temperature of primers smaller than 25 base pairs is the Wallace Rule ($T_d = 2(A+T) + 4(G+C)$). Computer programs can also be used to design primers, including but not limited to Array Designer Software (Arrayit Inc.), Oligonucleotide Probe Sequence Design Software for Genetic Analysis (Olympus Optical Co.), NetPrimer, and DNAsis from Hitachi Software Engineering. The TM (melting or annealing temperature) of each primer is calculated using software programs such as Net Primer (free web based program at

<http://premierbiosoft.com/netprimer/netpraunch/netpraunch.html>

(internet address as of February 13, 2002).

[0097] In another embodiment, the annealing temperature of the primers can be recalculated and increased after any cycle of amplification, including but not limited to cycle 1, 2, 3, 4, 5, cycles 6-10, cycles 10-15, cycles 15-20, cycles 20-25, cycles 25-30, cycles 30-35, or cycles 35-40. After the initial cycles of amplification, the 5' half of the primers is incorporated into the products from each loci of interest, thus the TM can be recalculated based on both the sequences of the 5' half and the 3' half of each primer.

[0098] For example, in FIG. 1B, the first cycle of amplification is performed at about the melting temperature of the 3' region of the second primer (region "c") that anneals to the template DNA, which is 13 bases. After the first cycle, the

annealing temperature can be raised to TM2, which is about the melting temperature of the 3' region of the first primer (region "b'") that anneals to the template DNA. The second primer cannot bind to the original template DNA because it only anneals to 13 bases in the original DNA template, and TM2 is about the melting temperature of approximately 20 bases, which is the 3' annealing region of the first primer (FIG. 1C). However, the first primer can bind to the DNA that was copied in the first cycle of the reaction. In the third cycle, the annealing temperature is raised to TM3, which is about the melting temperature of the entire sequence of the second primer ("c" and "d"). The template DNA produced from the second cycle of PCR contains both regions c' and d', and therefore, the second primer can anneal and extend at TM3 (FIG. 1D). The remaining cycles are performed at TM3. The entire sequence of the first primer (a + b') can anneal to the template from the third cycle of PCR, and extend (FIG. 1E). Increasing the annealing temperature will decrease non-specific binding and increase the specificity of the reaction, which is especially useful if amplifying a locus of interest from human genomic DNA, which contains 3×10^9 base pairs.

[0099] As used herein, the term "about" with regard to annealing temperatures is used to encompass temperatures within 10 degrees Celsius of the stated temperatures.

[0100] In one embodiment, one primer pair is used for each locus of interest. However, multiple primer pairs can be used for each locus of interest.

[0101] In one embodiment, primers are designed such that one or both primers of the primer pair contain sequence in the 5' region for one or more restriction endonucleases (restriction enzyme).

[0102] As used herein, with regard to the position at which restriction enzymes digest DNA, the "sense" strand is the strand reading 5' to 3' in the direction in which the restriction enzyme cuts. For example, BsmF I recognizes the following sequence:

5' GGGAC(N)₁₀¹ 3' (SEQ ID NO:1) or

3' CCCTG(N)₁₄ 5'
5' ↓(N)₁₄GTCCC 3' (SEQ ID NO:2)
3' ↑(N)₁₀CAGGG 5'

[0103] Thus, the sense strand is the strand containing the "GGGAC" sequence as it reads 5' to 3' in the direction that the restriction enzyme cuts.

[0104] As used herein, with regard to the position at which restriction enzymes digest DNA, the "antisense" strand is the strand reading 3' to 5' in the direction in which the restriction enzyme cuts. Thus, the antisense strand is the strand that contains the "ccctg" sequence as it reads 3' to 5'.

[0105] In the invention, one of the primers in a primer pair can be designed such that it contains a restriction enzyme recognition site for a restriction enzyme such that digestion with the restriction enzyme produces a recessed 3' end and a 5' overhang that contains the locus of interest (herein referred to as a "second primer"). For example, the second primer of a primer pair can contain a recognition site for a restriction enzyme that does not cut DNA at the recognition site but cuts "n" nucleotides away from the recognition site. "N" is a distance from the recognition site to the site of the cut by the restriction enzyme. If the recognition sequence is for the restriction enzyme BceA I, the enzyme will cut ten (10) nucleotides from the recognition site on the sense strand, and twelve (12) nucleotides away from the recognition site on the antisense strand.

[0106] The 3' region and preferably the 3' half of the primers is designed to anneal to a sequence that flanks the loci of interest (FIG. 1A). The second primer may anneal any distance from the locus of interest provided that digestion with the restriction enzyme that recognizes the restriction enzyme recognition site on this primer generates a 5' overhang that contains the locus of interest. The 5' overhang can be of any size, including but not limited to 1, 2, 3, 4, 5, 6, 7, 8, and more than 8 bases.

[0107] In a preferred embodiment, the 3' end of the second primer can anneal 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or more than 14 bases from the locus of interest or at the locus of interest.

[0108] In a preferred embodiment, the second primer is designed to anneal closer to the locus of interest than the other primer of a primer pair (the other primer is herein referred to as a "first primer"). The second primer can be a forward or reverse primer and the first primer can be a reverse or forward primer, respectively. Whether the first or second primer should be the forward or reverse primer can be determined by which design will provide better sequencing results.

[0109] For example, the primer that anneals closer to the locus of interest can contain a recognition site for the restriction enzyme BsmF I, which cuts ten (10) nucleotides from the recognition site on the sense strand, and fourteen (14) nucleotides from the recognition site on the antisense strand. In this case, the primer can be designed so that the restriction enzyme recognition site is 13 bases, 12 bases, 10 bases or 11 bases from the locus of interest. If the recognition site is 13 bases from the locus of interest, digestion with BsmF I will generate a 5' overhang (RXXX), wherein the locus of interest (R) is the first nucleotide in the overhang (reading 3' to 5'), and X is any nucleotide. If the recognition site is 12 bases from the locus of interest, digestion with BsmF I will generate a 5' overhang (XRXX), wherein the locus of interest (R) is the second nucleotide in the overhang (reading 3' to 5'). If the recognition site is 11 bases from the locus of interest, digestion with BsmF I will generate a 5' overhang (XXRX), wherein the locus of interest (R) is the third nucleotide in the overhang (reading 3' to 5'). The distance between the restriction enzyme recognition site and the locus of interest should be designed so that digestion with the restriction enzyme generates a 5' overhang, which contains the locus of interest. The effective distance between the recognition site and the locus of interest will vary depending on the choice of restriction enzyme.

[0110] In another embodiment, the second primer, which can anneal closer to the locus of interest relative to the first primer, can be designed so that the restriction enzyme that generates the 5' overhang, which contains the locus of interest, will see the same sequence at the cut site, independent of the nucleotide at the locus of interest. For example, if the primer that anneals closer to the locus

of interest is designed so that the recognition site for the restriction enzyme BsmF I (5' GGGAC 3') is thirteen bases from the locus of interest, the restriction enzyme will cut the antisense strand one base upstream of the locus of interest. The nucleotide at the locus of interest is adjacent to the cut site, and may vary from DNA molecule to DNA molecule. If it is desired that the nucleotides adjacent to the cut site be identical, the primer can be designed so that the restriction enzyme recognition site for BsmF I is twelve bases away from the locus of interest. Digestion with BsmF I will generate a 5' overhang, wherein the locus of interest is in the second position of the overhang (reading 3' to 5') and is no longer adjacent to the cut site. Designing the primer so that the restriction enzyme recognition site is twelve (12) bases from the locus of interest allows the nucleotides adjacent to the cut site to be the same, independent of the nucleotide at the locus of interest. Also, primers that have been designed so that the restriction enzyme recognition site is eleven (11) or ten (10) bases from the locus of interest will allow the nucleotides adjacent to the cut site to be the same, independent of the nucleotide at the locus of interest.

[0111] The 3' end of the first primer (either the forward or the reverse) can be designed to anneal at a chosen distance from the locus of interest. Preferably, for example, this distance is between 10-25, 25-50, 50-75, 75-100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, 450-500, 500-550, 550-600, 600-650, 650-700, 700-750, 750-800, 800-850, 850-900, 900-950, 950-1000 and greater than 1000 bases away from the locus of interest. The annealing sites of the first primers are chosen such that each successive upstream primer is further and further away from its respective downstream primer.

[0112] For example, if at locus of interest 1 the 3' ends of the first and second primers are Z bases apart, then at locus of interest 2, the 3' ends of the upstream and downstream primers are Z + K bases apart, where K = 1, 2, 3, 4, 5-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, or greater than 1000 bases (FIG 2). The purpose of making the upstream primers further and

further apart from their respective downstream primers is so that the PCR products of all the loci of interest differ in size and can be separated, e.g., on a sequencing gel. This allows for multiplexing by pooling the PCR products in later steps.

[0113] In one embodiment, the 5' region of the first primer can have a recognition site for any type of restriction enzyme. In a preferred embodiment, the first primer has at least one restriction enzyme recognition site that is different from the restriction enzyme recognition site in the second primer. In another preferred embodiment, the first primer anneals further away from the locus of interest than the second primer.

[0114] In a preferred embodiment, the second primer contains a restriction enzyme recognition sequence for a Type IIS restriction enzyme including but not limited to BceA I and BsmF I, which produce a two base 5' overhang and a four base 5' overhang, respectively. Restriction enzymes that are Type IIS are preferred because they recognize asymmetric base sequences (not palindromic like the orthodox Type II enzymes). Type IIS restriction enzymes cleave DNA at a specified position that is outside of the recognition site, typically up to 20 base pairs outside of the recognition site. These properties make Type IIS restriction enzymes, and the recognition sites thereof, especially useful in the method of the invention. Preferably, the Type IIS restriction enzymes used in this method leave a 5' overhang and a recessed 3' end.

[0115] A wide variety of Type IIS restriction enzymes are known and such enzymes have been isolated from bacteria, phage, archaebacteria and viruses of eukaryotic algae and are commercially available (Promega, Madison WI; New England Biolabs, Beverly, MA; Szybalski W. *et al.*, *Gene* 100:13-16, (1991)). Examples of Type IIS restriction enzymes that would be useful in the method of the invention include, but are not limited to enzymes such as those listed in Table I.

TABLE I:
TYPE IIS RESTRICTION ENZYMES THAT
GENERATE A 5' OVERHANG AND A RECESSED 3' END.

| Enzyme-Source | Recognition/ Cleavage Site | Supplier |
|--|-------------------------------|--|
| Alw I - <i>Acinetobacter lwoffii</i> | GGATC(4/5) | NE Biolabs |
| Alw26 I - <i>Acinetobacter lwoffii</i> | GTCTC(1/5) | Promega |
| Bbs I - <i>Bacillus laterosporus</i> | GAAGAC(2/6) | NE Biolabs |
| Bbv I - <i>Bacillus brevis</i> | GCAGC(8/12) | NE Biolabs |
| BceA I - <i>Bacillus cereus</i> 1315 | ACGGC(12/14) | NE Biolabs |
| Bmr I - <i>Bacillus megaterium</i> | ACTGGG(5/4) | NE Biolabs |
| Bsa I - <i>Bacillus stearothermophilus</i> 6-55 | GGTCTC(1/5) | NE Biolabs |
| Bst71 I - <i>Bacillus stearothermophilus</i> 71 | GCAGC(8/12) | Promega |
| BsmA I - <i>Bacillus stearothermophilus</i> A664 | GTCTC(1/5) | NE Biolabs |
| BsmB I - <i>Bacillus stearothermophilus</i> B61 | CGTCTC(1/5) | NE Biolabs |
| BsmF I - <i>Bacillus stearothermophilus</i> F | GGGAC(10/14) | NE Biolabs |
| BspM I - <i>Bacillus species</i> M | ACCTGC(4/8) | NE Biolabs |
| Ear I - <i>Enterobacter aerogenes</i> | CTCTTC(1/4) | NE Biolabs |
| Fau I - <i>Flavobacterium aquatile</i> | CCCGC(4/6) | NE Biolabs |
| Fok I - <i>Flavobacterium okeonokoites</i> | GGATG(9/13) | NE Biolabs |
| Hga I - <i>Haemophilus gallinarum</i> | GACGC(5/10) | NE Biolabs |
| Ple I - <i>Pseudomonas lemoignei</i> | GAGTC(4/5) | NE Biolabs |
| Sap I - <i>Saccharopolyspora species</i> | GCTCTTC(1/4) | NE Biolabs |
| SfaN I - <i>Streptococcus faecalis</i> ND547 | GCATC(5/9) | NE Biolabs |
| Sth132 I - <i>Streptococcus thermophilus</i> ST132 | CCCG(4/8) | No commercial supplier (Gene 195:201-206 (1997)) |

[0116] In one embodiment, a primer pair has sequence at the 5' region of each of the primers that provides a restriction enzyme recognition site that is unique for one restriction enzyme.

[0117] In another embodiment, a primer pair has sequence at the 5' region of each of the primers that provide a restriction site that is recognized by more than one restriction enzyme, and especially for more than one Type IIIS restriction enzyme. For example, certain consensus sequences can be recognized by more than one enzyme. For example, BsgI, Eco57I and BpmI all recognize the consensus 5' (G/C)TgnAG 3' and cleave 16 bp away on the antisense strand and 14 bp away on the sense strand. A primer that provides such a consensus sequence would result in a product that has a site that can be recognized by any of the restriction enzymes BsgI, Eco57I and BpmI.

[0118] Other restriction enzymes that cut DNA at a distance from the recognition site, and produce a recessed 3' end and a 5' overhang include Type III restriction enzymes. For example, the restriction enzyme EcoP15I recognizes the sequence 5' CAGCAG 3' and cleaves 25 bases downstream on the sense strand and 27 bases on the antisense strand. It will be further appreciated by a person of ordinary skill in the art that new restriction enzymes are continually being discovered and may readily be adopted for use in the subject invention.

[0119] In another embodiment, the second primer can contain a portion of the recognition sequence for a restriction enzyme, wherein the full recognition site for the restriction enzyme is generated upon amplification of the template DNA such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest. For example, the recognition site for BsmF I is 5' GGGACN₁₀ 3'. The 3' region, which anneals to the template DNA, of the second primer can end with the nucleotides "GGG," which do not have to be complementary with the template DNA. If the 3' annealing region is about 10-20 bases, even if the last three bases do not anneal, the primer will extend and, generate a BsmF I site.

Second primer: 5' GGAAATTCCATGATGCGTGGG→ (SEQ ID NO:3)

Template DNA: 3' CCTTAAGGTACTACGCAN₁·N₂·N₃·TG 5'

5' GGAAATTCCATGATGCGTN₁ N₂ N₃ AC 3' (SEQ ID NO:4)

[0120] The second primer can be designed to anneal to the template DNA, wherein the next two bases of the template DNA are thymidine and guanine, such

that an adenine and cytosine are incorporated into the primer forming a recognition site for BsmF I, 5' GGGACN₁₀ 3'. The second primer can be designed to anneal in such a manner that digestion with BsmF I generates a 5' overhang containing the locus of interest.

[0121] In another embodiment, the second primer can contain an entire or full recognition site for a restriction enzyme or a portion of a recognition site, which generates a full recognition site upon amplification of the template DNA such that digestion with a restriction enzyme that cuts at the recognition site generates a 5' overhang that contains the locus of interest. For example, the restriction enzyme BsaJ I binds the following recognition site: 5' C¹CN₁N₂GG 3'. The second primer can be designed such that the 3' region of the primer ends with "CC." The SNP of interest is represented by "N₁," and the template sequence downstream of the SNP is "N₂CC."

Second primer 5' GGAAATTCCATGATGCGTACC→ (SEQ ID NO:5)

Template DNA 3' CCTTAAGGTACTACGCATGGN₁N₂CC 5'

5' GGAAATTCCATGATGCGTACCN₁N₂GG 3' (SEQ ID

NO:6)

[0122] After digestion with BsaJ I, a 5' overhang of the following sequence would be generated:

5' C 3'
3' GGN₁N₂C 5'

[0123] If the nucleotide guanine is not reported at the locus of interest, the 3' recessed end can be filled in with unlabeled cytosine, which is complementary to the first nucleotide in the overhang. After removing the excess cytosine, labeled ddNTPs can be used to fill in the next nucleotide, N₁, which represents the locus of interest. Alternatively if guanine is reported to be a potential nucleotide at the locus of interest, labeled nucleotides can be used to detect a nucleotide 3' of the locus of interest. Unlabeled dCTP can be used to "fill in" followed by a fill in with a labeled nucleotide other than cytosine. Cytosine will be incorporated until it reaches a base that is not complementary. If the locus of interest contained a guanine, it would be filled in with the dCTP, which would allow incorporation of

the labeled nucleotide. However, if the locus of interest did not contain a guanine, the labeled nucleotide would not be incorporated. Other restriction enzymes can be used including but not limited to BssK I (5' \downarrow CCNGG 3'), Dde I (5' C \downarrow TNAG 3'), EcoN I (5' CCTNN \downarrow NNNAGG 3') (SEQ ID NO:7), Fnu4H I (5' GC \downarrow NGC 3'), Hinf I (5' G \downarrow ANTC 3'), PflF I (5' GACN \downarrow NNGTC 3'), Sau96 I (5' G \downarrow GNCC 3'), ScrF I (5' CC \downarrow NGG 3'), and Tth111 I (5' GACN \downarrow NNGTC 3').

[0124] It is not necessary that the 3' region, which anneals to the template DNA, of the second primer be 100% complementary to the template DNA. For example, the last 1, 2, or 3 nucleotides of the 3' end of the second primer can be mismatches with the template DNA. The region of the primer that anneals to the template DNA will target the primer, and allow the primer to extend. Even if, for example, the last two nucleotides are not complementary to the template DNA, the primer will extend and generate a restriction enzyme recognition site.

Second primer: 5' GGAAATTCCATGATGCGTACC→ (SEQ ID NO:5)

Template DNA: 3' CCTTTAAGGTACTACGCATN_aN_bN₁N₂CC 5'

5' GGAAATTCCATGATGCGTAN_aN_bN₁N₂GG 3' (SEQ ID NO:8)

[0125] After digestion with BsaJ I, a 5' overhang of the following sequence would be generated:

5' C 3'
3' GGN₁N₂C 5'

[0126] If the nucleotide cytosine is not reported at the locus of interest, the 5' overhang can be filled in with unlabeled cytosine. The excess cytosine can be rinsed away, and filled in with labeled ddNTPs. The first nucleotide incorporated (N₁) corresponds to the locus of interest.

[0127] Alternatively, it is possible to create the full restriction enzyme recognition sequence using the first and second primers. The recognition site for any restriction enzyme can be generated, as long as the recognition site contains at least one variable nucleotide. Restriction enzymes that recognize sites that

contain at least one variable nucleotide include but are not limited to BssK I (5'CCNGG 3'), Dde I (5'C[↓]TNAG 3'), Econ I (5'CCTNN[↓]NNNAGG 3') (SEQ ID NO:7), Fnu4H I (5'GC[↓]NGC 3'), Hinf I (5'G[↓]ANTC 3') PflF I (5' GACN[↓]NNGTC 3'), Sau96 I (5' G[↓]GNCC 3'), ScrF I (5' CC[↓]NGG 3'), and Tth111 I (5' GACN[↓]NNGTC 3'). In this embodiment, the first or second primer may anneal closer to the locus of interest or the first or second primer may anneal at an equal distance from the locus of interest. The first and second primers can be designed to contain mismatches to the template DNA at the 3' region; these mismatches create the restriction enzyme recognition site. The number of mismatches that can be tolerated at the 3' end depends on the length of the primer, and includes but is not limited to 1, 2, or more than 2 mismatches. For example, if the locus of interest is represented by N₁, a first primer can be designed to be complementary to the template DNA, depicted below as region "a." The 3' region of the first primer ends with "CC," which is not complementary to the template DNA. The second primer is designed to be complementary to the template DNA, which is depicted below as region "b". The 3' region of the second primer ends with "CC," which is not complementary to the template DNA.

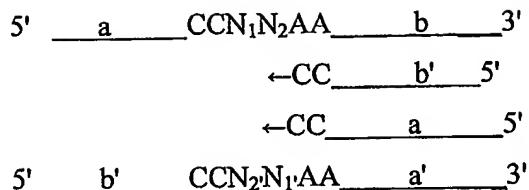
First primer 5' a CC→
 Template DNA 3' a' AAN₁N₂TT b' 5'
 5' a TTN₁N₂AA b 3'
 ←CC b' 5' Second Primer

[0128] After one round of amplification the following products would be generated:

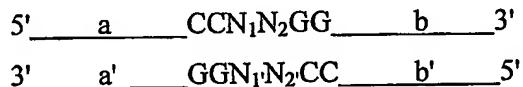
5' a CCN₁N₂AA b 3'
 and

5' b' CCN₂N₁AA a' 3'.

[0129] In cycle two, the primers can anneal to the templates that were generated from the first cycle of PCR:



[0130] After cycle two of PCR, the following products would be generated:



[0131] The restriction enzyme recognition site for BsaJ I is generated, and after digestion with BsaJ I, a 5' overhang containing the locus of interest is generated. The locus of interest can be detected as described in detail below. Alternatively, the 3' region of the first and second primers can contain 1, 2, 3, or more than 3 mismatches followed by a nucleotide that is complementary to the template DNA. For example, the first and second primers can be used to create a recognition site for the restriction enzyme EcoN I, which binds the following DNA sequence: 5' CCTNN¹NNNAGG 3'. The last nucleotides of each primer would be "CCTN₁ or CCTN₁N₂." The nucleotides "CCT" may or may not be complementary to the template DNA; however, N₁ and N₂ are nucleotides complementary to the template DNA. This allows the primers to anneal to the template DNA after the potential mismatches, which are used to create the restriction enzyme recognition site.

[0132] In another embodiment, a primer pair has sequence at the 5' region of each of the primers that provides two or more restriction sites that are recognized by two or more restriction enzymes.

[0133] In a most preferred embodiment, a primer pair has different restriction enzyme recognition sites at the 5' regions, especially 5' ends, such that a different restriction enzyme is required to cleave away any undesired sequences. For example, the first primer for locus of interest "A" can contain sequence recognized by a restriction enzyme, "X," which can be any type of restriction enzyme, and the second primer for locus of interest "A," which anneals closer to the locus of interest, can contain sequence for a restriction enzyme, "Y," which is

a Type IIS restriction enzyme that cuts "n" nucleotides away and leaves a 5' overhang and a recessed 3' end. The 5' overhang contains the locus of interest. After binding the amplified DNA to streptavidin coated wells, one can digest with enzyme "Y," rinse, then fill in with labeled nucleotides and rinse, and then digest with restriction enzyme "X," which will release the DNA fragment containing the locus of interest from the solid matrix. The locus of interest can be analyzed by detecting the labeled nucleotide that was "filled in" at the locus of interest, e.g. SNP site.

[0134] In another embodiment, the second primers for the different loci of interest that are being amplified according to the invention contain recognition sequence in the 5' regions for the same restriction enzyme and likewise all the first primers also contain the same restriction enzyme recognition site, which is a different enzyme from the enzyme that recognizes the second primers. The primer (either the forward or reverse primer) that anneals closer to the locus of interest contains a recognition site for, e.g., a Type IIs restriction enzyme.

[0135] In another embodiment, the second primers for the multiple loci of interest that are being amplified according to the invention contain restriction enzyme recognition sequences in the 5' regions for different restriction enzymes.

[0136] In another embodiment, the first primers for the multiple loci of interest that are being amplified according to the invention contain restriction enzyme recognition sequences in the 5' regions for different restriction enzymes.

[0137] Multiple restriction enzyme sequences provide an opportunity to influence the order in which pooled loci of interest are released from the solid support. For example, if 50 loci of interest are amplified, the first primers can have a tag at the extreme 5' end to aid in purification and a restriction enzyme recognition site, and the second primers can contain a recognition site for a type IIS restriction enzyme. For example, several of the first primers can have a restriction enzyme recognition site for EcoR I, other first primers can have a recognition site for Pst I, and still other first primers can have a recognition site for BamH I. After amplification, the loci of interest can be bound to a solid

support with the aid of the tag on the first primers. By performing the restriction digests one restriction enzyme at a time, one can serially release the amplified loci of interest. If the first digest is performed with EcoRI, the loci of interest amplified with the first primers containing the recognition site for EcoRI will be released, and collected while the other loci of interest remain bound to the solid support. The amplified loci of interest can be selectively released from the solid support by digesting with one restriction enzyme at a time. The use of different restriction enzyme recognition sites in the first primers allows a larger number of loci of interest to be amplified in a single reaction tube.

[0138] In a preferred embodiment, any region 5' of the restriction enzyme digestion site of each primer can be modified with a functional group that provides for fragment manipulation, processing, identification, and/or purification. Examples of such functional groups, or tags, include but are not limited to biotin, derivatives of biotin, carbohydrates, haptens, dyes, radioactive molecules, antibodies, and fragments of antibodies, peptides, and immunogenic molecules.

[0139] In another embodiment, the template DNA can be replicated once, without being amplified beyond a single round of replication. This is useful when there is a large amount of the DNA available for analysis such that a large number of copies of the loci of interest are already present in the sample, and further copies are not needed. In this embodiment, the primers are preferably designed to contain a "hairpin" structure in the 5' region, such that the sequence doubles back and anneals to a sequence internal to itself in a complementary manner. When the template DNA is replicated only once, the DNA sequence comprising the recognition site would be single-stranded if not for the "hairpin" structure. However, in the presence of the hairpin structure, that region is effectively double stranded, thus providing a double stranded substrate for activity by restriction enzymes.

[0140] To the extent that the reaction conditions are compatible, all the primer pairs to analyze a locus or loci of interest of DNA can be mixed together for use

in the method of the invention. In a preferred embodiment, all primer pairs are mixed with the template DNA in a single reaction vessel. Such a reaction vessel can be, for example, a reaction tube, or a well of a microtiter plate.

[0141] Alternatively, to avoid competition for nucleotides and to minimize primer dimers and difficulties with annealing temperatures for primers, each locus of interest or small groups of loci of interest can be amplified in separate reaction tubes or wells, and the products later pooled if desired. For example, the separate reactions can be pooled into a single reaction vessel before digestion with the restriction enzyme that generates a 5' overhang, which contains the locus of interest or SNP site, and a 3' recessed end. Preferably, the primers of each primer pair are provided in equimolar amounts. Also, especially preferably, each of the different primer pairs is provided in equimolar amounts relative to the other pairs that are being used.

[0142] In another embodiment, combinations of primer pairs that allow efficient amplification of their respective loci of interest can be used (see e.g. FIG. 2). Such combinations can be determined prior to use in the method of the invention. Multi-well plates and PCR machines can be used to select primer pairs that work efficiently with one another. For example, gradient PCR machines, such as the Eppendorf Mastercycler® gradient PCR machine, can be used to select the optimal annealing temperature for each primer pair. Primer pairs that have similar properties can be used together in a single reaction tube.

[0143] In another embodiment, a multi-sample container including but not limited to a 96-well or more plate can be used to amplify a single locus of interest with the same primer pairs from multiple template DNA samples with optimal PCR conditions for that locus of interest. Alternatively, a separate multi-sample container can be used for amplification of each locus of interest and the products for each template DNA sample later pooled. For example, gene A from 96 different DNA samples can be amplified in microtiter plate 1, gene B from 96 different DNA samples can be amplified in microtiter plate 2, etc., and then the amplification products can be pooled.

[0144] The result of amplifying multiple loci of interest is a preparation that contains representative PCR products having the sequence of each locus of interest. For example, if DNA from only one individual is used as the template DNA and if hundreds of disease-related loci of interest were amplified from the template DNA, the amplified DNA would be a mixture of small, PCR products from each of the loci of interest. Such a preparation could be further analyzed at that time to determine the sequence at each locus of interest or at only some of loci of interest. Additionally, the preparation could be stored in a manner that preserves the DNA and can be analyzed at a later time. Information contained in the amplified DNA can be revealed by any suitable method including but not limited to fluorescence detection, sequencing, gel electrophoresis, and mass spectrometry (see "Detection of Incorporated Nucleotide" section below).

Amplification of Loci of Interest

[0145] The template DNA can be amplified using any suitable method known in the art including but not limited to PCR (polymerase chain reaction), 3SR (self-sustained sequence reaction), LCR (ligase chain reaction), RACE-PCR (rapid amplification of cDNA ends), PLCR (a combination of polymerase chain reaction and ligase chain reaction), Q-beta phage amplification (Shah et al., *J. Medical Micro.* 33: 1435-41 (1995)), SDA (strand displacement amplification), SOE-PCR (splice overlap extension PCR), and the like. These methods can be used to design variations of the releasable primer mediated cyclic amplification reaction explicitly described in this application. In the most preferred embodiment, the template DNA is amplified using PCR (*PCR: A Practical Approach*, M. J. McPherson, et al., IRL Press (1991); *PCR Protocols: A Guide to Methods and Applications*, Innis, et al., Academic Press (1990); and *PCR Technology: Principles and Applications of DNA Amplification*, H. A. Erlich, Stockton Press (1989)). PCR is also described in numerous U.S. patents, including U.S. Pat. Nos. 4,683,195; 4,683,202; 4,800,159; 4,965,188; 4,889,818; 5,075,216; 5,079,352; 5,104,792, 5,023,171; 5,091,310; and 5,066,584.

[0146] The components of a typical PCR reaction include but are not limited to a template DNA, primers, a reaction buffer (dependent on choice of polymerase), dNTPs (dATP, dTTP, dGTP, and dCTP) and a DNA polymerase. Suitable PCR primers can be designed and prepared as discussed above (see "Primer Design" section above). Briefly, the reaction is heated to 95°C for 2 min. to separate the strands of the template DNA, the reaction is cooled to an appropriate temperature (determined by calculating the annealing temperature of designed primers) to allow primers to anneal to the template DNA, and heated to 72°C for two minutes to allow extension.

[0147] In a preferred embodiment, the annealing temperature is increased in each of the first three cycles of amplification to reduce non-specific amplification. See also Example 1, below. The TM1 of the first cycle of PCR is about the melting temperature of the 3' region of the second primer that anneals to the template DNA. The annealing temperature can be raised in cycles 2-10, preferably in cycle 2, to TM2, which is about the melting temperature of the 3' region, which anneals to the template DNA, of the first primer. If the annealing temperature is raised in cycle 2, the annealing temperature remains about the same until the next increase in annealing temperature. Finally, in any cycle subsequent to the cycle in which the annealing temperature was increased to TM2, preferably cycle 3, the annealing temperature is raised to TM3, which is about the melting temperature of the entire second primer. After the third cycle, the annealing temperature for the remaining cycles may be at about TM3 or may be further increased. In this example, the annealing temperature is increased in cycles 2 and 3. However, the annealing temperature can be increased from a low annealing temperature in cycle 1 to a high annealing temperature in cycle 2 without any further increases in temperature or the annealing temperature can progressively change from a low annealing temperature to a high annealing temperature in any number of incremental steps. For example, the annealing temperature can be changed in cycles 2, 3, 4, 5, 6, etc.

[0148] After annealing, the temperature in each cycle is increased to an "extension" temperature to allow the primers to "extend" and then following extension the temperature in each cycle is increased to the denaturization temperature. For PCR products less than 500 base pairs in size, one can eliminate the extension step in each cycle and just have denaturization and annealing steps. A typical PCR reaction consists of 25-45 cycles of denaturation, annealing and extension as described above. However, as previously noted, even only one cycle of amplification (one copy) can be sufficient for practicing the invention.

[0149] Any DNA polymerase that catalyzes primer extension can be used including but not limited to E. coli DNA polymerase, Klenow fragment of E. coli DNA polymerase I, T7 DNA polymerase, T4 DNA polymerase, Taq polymerase, Pfu DNA polymerase, Vent DNA polymerase, bacteriophage 29, and REDTaq™ Genomic DNA polymerase, or sequenase. Preferably, a thermostable DNA polymerase is used. A "hot start" PCR can also be performed wherein the reaction is heated to 95°C for two minutes prior to addition of the polymerase or the polymerase can be kept inactive until the first heating step in cycle 1. "Hot start" PCR can be used to minimize nonspecific amplification. Any number of PCR cycles can be used to amplify the DNA, including but not limited to 2, 5, 10, 15, 20, 25, 30, 35, 40, or 45 cycles. In a most preferred embodiment, the number of PCR cycles performed is such that equimolar amounts of each loci of interest are produced.

Purification of Amplified DNA

[0150] Purification of the amplified DNA is not necessary for practicing the invention. However, in one embodiment, if purification is preferred, the 5' end of the primer (first or second primer) can be modified with a tag that facilitates purification of the PCR products. In a preferred embodiment, the first primer is modified with a tag that facilitates purification of the PCR products. The modification is preferably the same for all primers, although different modifications can be used if it is desired to separate the PCR products into different groups.

[0151] The tag can be a radioisotope, fluorescent reporter molecule, chemiluminescent reporter molecule, antibody, antibody fragment, hapten, biotin, derivative of biotin, photobiotin, iminobiotin, digoxigenin, avidin, enzyme, acridinium, sugar, enzyme, apoenzyme, homopolymeric oligonucleotide, hormone, ferromagnetic moiety, paramagnetic moiety, diamagnetic moiety, phosphorescent moiety, luminescent moiety, electrochemiluminescent moiety, chromatic moiety, moiety having a detectable electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity, or combinations thereof.

[0152] In a preferred embodiment, the 5' ends of the primers can be biotinylated (Kandpal et al., Nucleic Acids Res. 18:1789-1795 (1990); Kaneoka et al., Biotechniques 10:30-34 (1991); Green et al., Nucleic Acids Res. 18:6163-6164 (1990)). The biotin provides an affinity tag that can be used to purify the copied DNA from the genomic DNA or any other DNA molecules that are not of interest. Biotinylated molecules can be purified using a streptavidin coated matrix as shown in FIG. 1F, including but not limited to Streptawell, transparent, High-Bind plates from Roche Molecular Biochemicals (catalog number 1 645 692, as listed in Roche Molecular Biochemicals, 2001 Biochemicals Catalog).

[0153] The PCR product of each locus of interest is placed into separate wells of a Streptavidin coated plate. Alternatively, the PCR products of the loci of interest can be pooled and placed into a streptavidin coated matrix, including but not limited to the Streptawell, transparent, High-Bind plates from Roche Molecular Biochemicals (catalog number 1 645 692, as listed in Roche Molecular Biochemicals, 2001 Biochemicals Catalog).

[0154] The amplified DNA can also be separated from the template DNA using non-affinity methods known in the art, for example, by polyacrylamide gel electrophoresis using standard protocols.

Digestion of Amplified DNA

[0155] The amplified DNA can be digested with a restriction enzyme that recognizes a sequence that had been provided on the first or second primer using standard protocols known within the art (FIGS. 6A-6D). The enzyme used

depends on the restriction recognition site generated with the first or second primer. See "Primer Design" section, above, for details on restriction recognition sites generated on primers.

[0156] Type IIS restriction enzymes are extremely useful in that they cut approximately 10-20 base pairs outside of the recognition site. Preferably, the Type IIS restriction enzymes used are those that generate a 5' overhang and a recessed 3' end, including but not limited to BceA I and BsmF I (see e.g. Table I). In a most preferred embodiment, the second primer (either forward or reverse), which anneals close to the locus of interest, contains a restriction enzyme recognition sequence for BsmF I or BceA I. The Type IIS restriction enzyme BsmF I recognizes the nucleic acid sequence GGGAC, and cuts 14 nucleotides from the recognition site on the antisense strand and 10 nucleotides from the recognition site on the sense strand. Digestion with BsmF I generates a 5' overhang of four (4) bases.

[0157] For example, if the second primer is designed so that after amplification the restriction enzyme recognition site is 13 bases from the locus of interest, then after digestion, the locus of interest is the first base in the 5' overhang (reading 3' to 5'), and the recessed 3' end is one base upstream of the locus of interest. The 3' recessed end can be filled in with a nucleotide that is complementary to the locus of interest. One base of the overhang can be filled in using dideoxynucleotides. However, 1, 2, 3, or all 4 bases of the overhang can be filled in using deoxynucleotides or a mixture of dideoxynucleotides and deoxynucleotides.

[0158] The restriction enzyme BsmF I cuts DNA ten (10) nucleotides from the recognition site on the sense strand and fourteen (14) nucleotides from the recognition site on the antisense strand. However, in a sequence dependent manner, the restriction enzyme BsmF I also cuts eleven (11) nucleotides from the recognition site on the sense strand and fifteen (15) nucleotides from the recognition site on the antisense strand. Thus, two populations of DNA molecules exist after digestion: DNA molecules cut at 10/14 and DNA molecules cut at

11/15. If the recognition site for BsmF I is 13 bases from the locus of interest in the amplified product, then DNA molecules cut at the 11/15 position will generate a 5' overhang that contains the locus of interest in the second position of the overhang (reading 3' to 5'). The 3' recessed end of the DNA molecules can be filled in with labeled nucleotides. For example, if labeled dideoxynucleotides are used, the 3' recessed end of the molecules cut at 11/15 would be filled in with one base, which corresponds to the base upstream of the locus of interest, and the 3' recessed end of molecules cut at 10/14 would be filled in with one base, which corresponds to the locus of interest. The DNA molecules that have been cut at the 10/14 position and the DNA molecules that have been cut at the 11/15 position can be separated by size, and the incorporated nucleotides detected. This allows detection of both the nucleotide before the locus of interest, detection of the locus of interest, and potentially the three bases pairs after the locus of interest.

[0159] Alternatively, if the base upstream of the locus of interest and the locus of interest are different nucleotides, then the 3' recessed end of the molecules cut at 11/15 can be filled in with deoxynucleotide that is complementary to the upstream base. The remaining deoxynucleotide is washed away, and the locus of interest site can be filled in with either labeled deoxynucleotides, unlabeled deoxynucleotides, labeled dideoxynucleotides, or unlabeled dideoxynucleotides. After the fill in reaction, the nucleotide can be detected by any suitable method. Thus, after the first fill in reaction with dNTP, the 3' recessed end of the molecules cut at 10/14 and 11/15 is upstream of the locus of interest. The 3' recessed end can now be filled in one base, which corresponds to the locus of interest, two bases, three bases or four bases.

[0160] Alternatively, if the base upstream of the locus of interest and the base downstream of the locus of interest are reported to be the same, the 3' recessed end of the molecules cut at 11/15 can be "filled in" with unlabeled deoxynucleotide, followed by a "fill in" with labeled dideoxynucleotide. For example, if the nucleotide upstream of the locus of interest is a cytosine, and a cytosine is a potential nucleotide at the locus of interest, and an adenine is the

first nucleotide 3' of the locus of interest, a "fill in" reaction can be performed with unlabeled deoxyguanine triphosphate (dGTP), followed by a fill in with labeled dideoxythymidine triphosphate. If the locus of interest contains a cytosine, the ddTTP will be incorporated and detected. However, if the locus of interest does not contain a cytosine, the dGTP will not be incorporated, which prevents incorporation of the ddTTP.

[0161] The restriction enzyme BceA I recognizes the nucleic acid sequence ACGGC and cuts 12 (twelve) nucleotides from the recognition site on the sense strand and 14 (fourteen) nucleotides from the recognition site on the antisense strand. If the distance from the recognition site for BceA I on the second primer is designed to be thirteen (13) bases from the locus of interest (see FIGS. 4A-4D), digestion with BceA I will generate a 5' overhang of two bases, which contains the locus of interest, and a recessed 3' end that is upstream of the locus of interest. The locus of interest is the first nucleotide in the 5' overhang (reading 3' to 5').

[0162] Alternative cutting is also seen with the restriction enzyme BceA I, although at a much lower frequency than is seen with BsmF I. The restriction enzyme BceA I can cut thirteen (13) nucleotides from the recognition site on the sense strand and fifteen (15) nucleotides from the recognition site on the antisense strand. Thus, two populations of DNA molecules exist: DNA molecules cut at 12/14 and DNA molecules cut at 13/15. If the restriction enzyme recognition site is 13 bases from the locus of interest in the amplified product, DNA molecules cut at the 13/15 position yield a 5' overhang, which contains the locus of interest in the second position of the overhang (reading 3' to 5'). Labeled dideoxynucleotides can be used to fill in the 3' recessed end of the DNA molecules. The DNA molecules cut at 13/15 will have the base upstream of the locus of interest filled in, and the DNA molecules cut at 12/14 will have the locus of interest site filled in. The DNA molecules cut at 13/15 and those cut at 12/14 can be separated by size, and the incorporated nucleotide detected. Thus, the alternative cutting can be used to obtain additional sequence information.

[0163] Alternatively, if the two bases in the 5' overhang are different, the 3' recessed end of the DNA molecules, which were cut at 13/15, can be filled in with the deoxynucleotide complementary to the first base in the overhang, and excess deoxynucleotide washed away. After filling in, the 3' recessed end of the DNA molecules that were cut at 12/14 and the DNA molecules that were cut at 13/15 are upstream of the locus of interest. The 3' recessed ends can be filled with either labeled dideoxynucleotides, unlabeled dideoxynucleotides, labeled deoxynucleotides, or unlabeled deoxynucleotides.

[0164] If the primers provide different restriction sites for certain of the loci of interest that were copied, all the necessary restriction enzymes can be added together to digest the copied DNA simultaneously. Alternatively, the different restriction digests can be made in sequence, for example, using one restriction enzyme at a time, so that only the product that is specific for that restriction enzyme is digested.

Incorporation of Labeled Nucleotides

[0165] Digestion with the restriction enzyme that recognizes the sequence on the second primer generates a recessed 3' end and a 5' overhang, which contains the locus of interest (FIG. 1G). The recessed 3' end can be filled in using the 5' overhang as a template in the presence of unlabeled or labeled nucleotides or a combination of both unlabeled and labeled nucleotides. The nucleotides can be labeled with any type of chemical group or moiety that allows for detection including but not limited to radioactive molecules, fluorescent molecules, antibodies, antibody fragments, haptens, carbohydrates, biotin, derivatives of biotin, phosphorescent moieties, luminescent moieties, electrochemiluminescent moieties, chromatic moieties, and moieties having a detectable electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity. The nucleotides can be labeled with one or more than one type of chemical group or moiety. Each nucleotide can be labeled with the same chemical group or moiety. Alternatively, each different nucleotide can be labeled with a different chemical group or moiety. The labeled nucleotides can be dNTPs, ddNTPs, or a

mixture of both dNTPs and ddNTPs. The unlabeled nucleotides can be dNTPs, ddNTPs or a mixture of both dNTPs and ddNTPs.

[0166] Any combination of nucleotides can be used to incorporate nucleotides including but not limited to unlabeled deoxynucleotides, labeled deoxynucleotides, unlabeled dideoxynucleotides, labeled dideoxynucleotides, a mixture of labeled and unlabeled deoxynucleotides, a mixture of labeled and unlabeled dideoxynucleotides, a mixture of labeled deoxynucleotides and labeled dideoxynucleotides, a mixture of labeled deoxynucleotides and unlabeled dideoxynucleotides, a mixture of unlabeled deoxynucleotides and unlabeled dideoxynucleotides, a mixture of unlabeled deoxynucleotides and labeled dideoxynucleotides, dideoxynucleotide analogues, deoxynucleotide analogues, a mixture of dideoxynucleotide analogues and deoxynucleotide analogues, phosphorylated nucleoside analogues, 2-deoxynucleoside-5' triphosphates and modified 2'-deoxynucleoside triphosphates.

[0167] For example, as shown in FIG. 1H, in the presence of a polymerase, the 3' recessed end can be filled in with fluorescent ddNTP using the 5' overhang as a template. The incorporated ddNTP can be detected using any suitable method including but not limited to fluorescence detection.

[0168] All four nucleotides can be labeled with different fluorescent groups, which will allow one reaction to be performed in the presence of all four labeled nucleotides. Alternatively, five separate "fill in" reactions can be performed for each locus of interest; each of the four reactions will contain a different labeled nucleotide (e.g. ddATP*, ddTTP*, ddUTP*, ddGTP*, or ddCTP*, where * indicates a labeled nucleotide). Each nucleotide can be labeled with different chemical groups or the same chemical groups. The labeled nucleotides can be dideoxynucleotides or deoxynucleotides.

[0169] In another embodiment, nucleotides can be labeled with fluorescent dyes including but not limited to fluorescein, pyrene, 7-methoxycoumarin, Cascade Blue.TM., Alexa Flur 350, Alexa Flur 430, Alexa Flur 488, Alexa Flur 532, Alexa Flur 546, Alexa Flur 568, Alexa Flur 594, Alexa Flur 633, Alexa Flur

647, Alexa Flur 660, Alexa Flur 680, AMCA-X, dialkylaminocoumarin, Pacific Blue, Marina Blue, BODIPY 493/503, BODIPY Fl-X, DTAF, Oregon Green 500, Dansyl-X, 6-FAM, Oregon Green 488, Oregon Green 514, Rhodamine Green-X, Rhodol Green, Calcein, Eosin, ethidium bromide, NBD, TET, 2', 4', 5', 7' tetrabromosulfonefluorescien, BODIPY-R6G, BODIPY-Fl BR2, BODIPY 530/550, HEX, BODIPY 558/568, BODIPY-TMR-X., PyMPO, BODIPY 564/570, TAMRA, BODIPY 576/589, Cy3, Rhodamine Red-x, BODIPY 581/591, carboxyXrhodamine, Texas Red-X, BODIPY-TR-X., Cy5, SpectrumAqua, SpectrumGreen #1, SpectrumGreen #2, SpectrumOrange, SpectrumRed, or naphthofluorescein.

[0170] In another embodiment, the "fill in" reaction can be performed with fluorescently labeled dNTPs, wherein the nucleotides are labeled with different fluorescent groups. The incorporated nucleotides can be detected by any suitable method including but not limited to Fluorescence Resonance Energy Transfer (FRET).

[0171] In another embodiment, a mixture of both labeled ddNTPs and unlabeled dNTPs can be used for filling in the recessed 3' end of the DNA sequence containing the SNP or locus of interest. Preferably, the 5' overhang consists of more than one base, including but not limited to 2, 3, 4, 5, 6 or more than 6 bases. For example, if the 5' overhang consists of the sequence "XGAA," wherein X is the locus of interest, e.g. SNP, then filling in with a mixture of labeled ddNTPs and unlabeled dNTPs will produce several different DNA fragments. If a labeled ddNTP is incorporated at position "X," the reaction will terminate and a single labeled base will be incorporated. If however, an unlabeled dNTP is incorporated, the polymerase continues to incorporate other bases until a labeled ddNTP is incorporated. If the first two nucleotides incorporated are dNTPs, and the third is a ddNTP, the 3' recessed end will be extended by three bases. This DNA fragment can be separated from the other DNA fragments that were extended by 1, 2, or 4 bases by size. A mixture of labeled ddNTPs and unlabeled dNTPs will allow all bases of the overhang to be filled in, and provides

additional sequence information about the locus of interest, e.g. SNP (see FIGS. 7E and 9D).

[0172] After incorporation of the labeled nucleotide, the amplified DNA can be digested with a restriction enzyme that recognizes the sequence provided by the first primer. For example, in FIG 1I, the amplified DNA is digested with a restriction enzyme that binds to region "a," which releases the DNA fragment containing the incorporated nucleotide from the streptavidin matrix.

[0173] Alternatively, one primer of each primer pair for each locus of interest can be attached to a solid support matrix including but not limited to a well of a microtiter plate. For example, streptavidin-coated microtiter plates can be used for the amplification reaction with a primer pair, wherein one primer is biotinylated. First, biotinylated primers are bound to the streptavidin-coated microtiter plates. Then, the plates are used as the reaction vessel for PCR amplification of the loci of interest. After the amplification reaction is complete, the excess primers, salts, and template DNA can be removed by washing. The amplified DNA remains attached to the microtiter plate. The amplified DNA can be digested with a restriction enzyme that recognizes a sequence on the second primer and generates a 5' overhang, which contains the locus of interest. The digested fragments can be removed by washing. After digestion, the SNP site or locus of interest is exposed in the 5' overhang. The recessed 3' end is filled in with a labeled nucleotide, including but not limited to, fluorescent ddNTP in the presence of a polymerase. The labeled DNA can be released into the supernatant in the microtiter plate by digesting with a restriction enzyme that recognizes a sequence in the 5' region of the first primer.

Analysis of the locus of interest

[0174] The labeled loci of interest can be analyzed by a variety of methods including but not limited to fluorescence detection, DNA sequencing gel, capillary electrophoresis on an automated DNA sequencing machine, microchannel electrophoresis, and other methods of sequencing, mass spectrometry, time of flight mass spectrometry, quadrupole mass spectrometry, magnetic sector mass

spectrometry, electric sector mass spectrometry infrared spectrometry, ultraviolet spectrometry, palentiostatic amperometry or by DNA hybridization techniques including Southern Blots, Slot Blots, Dot Blots, and DNA microarrays, wherein DNA fragments would be useful as both "probes" and "targets," ELISA, fluorimetry, and Fluorescence Resonance Energy Transfer (FRET).

[0175] The loci of interest can be analyzed using gel electrophoresis followed by fluorescence detection of the incorporated nucleotide. Another method to analyze or read the loci of interest is to use a fluorescent plate reader or fluorimeter directly on the 96-well streptavidin coated plates. The plate can be placed onto a fluorescent plate reader or scanner such as the Pharmacia 9200 Typhoon to read each locus of interest.

[0176] Alternatively, the PCR products of the loci of interest can be pooled and after "filling in," (FIG. 10) the products can be separated by size, using any method appropriate for the same, and then analyzed using a variety of techniques including but not limited to fluorescence detection, DNA sequencing gel, capillary electrophoresis on an automated DNA sequencing machine, microchannel electrophoresis, other methods of sequencing, DNA hybridization techniques including Southern Blots, Slot Blots, Dot Blots, and DNA microarrays, mass spectrometry, time of flight mass spectrometry, quadrupole mass spectrometry, magnetic sector mass spectrometry, electric sector mass spectrometry infrared spectrometry, ultraviolet spectrometry, palentiostatic amperometry. For example, polyacrylamide gel electrophoresis can be used to separate DNA by size and the gel can be scanned to determine the color of fluorescence in each band (using e.g. ABI 377 DNA sequencing machine or a Pharmacia Typhoon 9200).

[0177] In another embodiment, one nucleotide can be used to determine the sequence of multiple alleles of a gene. A nucleotide that terminates the elongation reaction can be used to determine the sequence of multiple alleles of a gene. At one allele, the terminating nucleotide is complementary to the locus of interest in the 5' overhang of said allele. The nucleotide is incorporated and terminates the reaction. At a different allele, the terminating nucleotide is not complementary to

the locus of interest, which allows a non-terminating nucleotide to be incorporated at the locus of interest of the different allele. However, the terminating nucleotide is complementary to a nucleotide downstream from the locus of interest in the 5' overhang of said different allele. The sequence of the alleles can be determined by analyzing the patterns of incorporation of the terminating nucleotide. The terminating nucleotide can be labeled or unlabeled.

[0178] In another embodiment, the terminating nucleotide is a nucleotide that terminates or hinders the elongation reaction including but not limited to a dideoxynucleotide, a dideoxynucleotide derivative, a dideoxynucleotide analog, a dideoxynucleotide homolog, a dideoxynucleotide with a sulfur chemical group, a deoxynucleotide, a deoxynucleotide derivative, a deoxynucleotide homolog, a deoxynucleotide analog, and a deoxynucleotide with a sulfur chemical group, arabinoside triphosphate, an arabinoside triphosphate analog, a arabinoside triphosphate homolog, or an arabinoside derivative.

[0179] In another embodiment, a terminating nucleotide labeled with one signal generating moiety tag, including but not limited to a fluorescent dye, can be used to determine the sequence of the alleles of a locus of interest. The use of a single nucleotide labeled with one signal generating moiety tag eliminates any difficulties that can arise when using different fluorescent moieties. In addition, using one nucleotide labeled with one signal generating moiety tag to determine the sequence of alleles of a locus of interest reduces the number of reactions, and eliminates pipetting errors.

[0180] For example, if the second primer contains the restriction enzyme recognition site for BsmFI, digestion will generate a 5' overhang of 4 bases. The second primer can be designed such that the locus of interest is located in the first position of the overhang. A representative overhang is depicted below, where R represents the locus of interest:

| | | | | | |
|-------------------|---|---|---|---|--|
| 5' CAC | | | | | |
| 3' GTG | R | T | G | G | |
| Overhang position | 1 | 2 | 3 | 4 | |

[0181] One nucleotide with one signal generating moiety tag can be used to determine whether the variable site is homozygous or heterozygous. For example, if the variable site is adenine (A) or guanine (G), then either adenine or guanine can be used to determine the sequence of the alleles of the locus of interest, provided that there is an adenine or guanine in the overhang at position 2, 3, or 4.

[0182] For example, if the nucleotide in position 2 of the overhang is thymidine, which is complementary to adenine, then labeled ddATP, unlabeled dCTP, dGTP, and dTTP can be used to determine the sequence of the alleles of the locus of interest. The ddATP can be labeled with any signal generating moiety including but not limited to a fluorescent dye. If the template DNA is homozygous for adenine, then labeled ddATP* will be incorporated at position 1 complementary to the overhang at the alleles, and no nucleotide incorporation will be seen at position 2, 3 or 4 complementary to the overhang.

| | | | | | |
|-------------------|--------|----|---|---|---|
| Allele 1 | 5' CCC | A* | | | |
| | 3' GGG | T | T | G | G |
| Overhang position | | 1 | 2 | 3 | 4 |
| Allele 2 | 5' CCC | A* | | | |
| | 3' GGG | T | T | G | G |
| Overhang position | | 1 | 2 | 3 | 4 |

[0183] One signal will be seen corresponding to incorporation of labeled ddATP at position 1 complementary to the overhang, which indicates that the individual is homozygous for adenine at this position. This method of labeling eliminates any difficulties that may arise from using different dyes that have different quantum coefficients.

Homozygous guanine:

[0184] If the template DNA is homozygous for guanine, then no ddATP will be incorporated at position 1 complementary to the overhang, but ddATP will be incorporated at the first available position, which in this case is position 2

complementary to the overhang. For example, if the second position in the overhang corresponds to a thymidine, then:

| | | | | | |
|-------------------|--------|----------|-----------|---|---|
| Allele 1 | 5' CCC | G | A* | | |
| | 3' GGG | C | T | G | G |
| Overhang position | | 1 | 2 | 3 | 4 |
| Allele 2 | 5' CCC | G | A* | | |
| | 3' GGG | C | T | G | G |
| Overhang position | | 1 | 2 | 3 | 4 |

[0185] One signal will be seen corresponding to incorporation of ddATP at position 2 complementary to the overhang, which indicates that the individual is homozygous for guanine. The molecules that are filled in at position 2 complementary to the overhang will have a different molecular weight than the molecules filled in at position 1 complementary to the overhang.

[0186] Heterozygous condition:

| | | | | | |
|-------------------|--------|-----------|-----------|---|---|
| Allele 1 | 5' CCC | A* | | | |
| | 3' GGG | T | T | G | G |
| Overhang position | | 1 | 2 | 3 | 4 |
| Allele 2 | 5' CCC | G | A* | | |
| | 3' GGG | C | T | G | G |
| Overhang position | | 1 | 2 | 3 | 4 |

[0187] Two signals will be seen; the first signal corresponds to the ddATP filled in at position one complementary to the overhang and the second signal corresponds to the ddATP filled in at position 2 complementary to the overhang. The two signals can be separated based on molecular weight; allele 1 and allele 2 will be separated by a single base pair, which allows easy detection and quantitation of the signals. Molecules filled in at position one can be

distinguished from molecules filled in at position two using any method that discriminates based on molecular weight including but not limited to gel electrophoresis, capillary gel electrophoresis, DNA sequencing, and mass spectrometry. It is not necessary that the nucleotide be labeled with a chemical moiety; the DNA molecules corresponding to the different alleles can be separated based on molecular weight.

[0188] If position 2 of the overhang is not complementary to adenine, it is possible that positions 3 or 4 may be complementary to adenine. For example, position 3 of the overhang may be complementary to the nucleotide adenine, in which case labeled ddATP may be used to determine the sequence of both alleles.

[0189] Homozygous for adenine:

| | | | | | |
|-------------------|--------|----|---|---|---|
| Allele 1 | 5' CCC | A* | | | |
| | 3' GGG | T | G | T | G |
| Overhang position | | 1 | 2 | 3 | 4 |
| Allele 2 | 5' CCC | A* | | | |
| | 3' GGG | T | G | T | G |
| Overhang position | | 1 | 2 | 3 | 4 |

[0190] Homozygous for guanine:

| | | | | | |
|-------------------|--------|---|---|----|---|
| Allele 1 | 5' CCC | G | C | A* | |
| | 3' GGG | C | G | T | G |
| Overhang position | | 1 | 2 | 3 | 4 |
| Allele 2 | 5' CCC | G | C | A* | |
| | 3' GGG | C | G | T | G |
| Overhang position | | 1 | 2 | 3 | 4 |

[0191] Heterozygous:

| | | | | | |
|-------------------|--------|----|---|---|---|
| Allele 1 | 5' CCC | A* | | | |
| | 3' GGG | T | G | T | G |
| Overhang position | | 1 | 2 | 3 | 4 |

| | | | | | |
|-------------------|--------|---|---|----|---|
| Allele 2 | 5' CCC | G | C | A* | |
| | 3' GGG | C | G | T | G |
| Overhang position | | 1 | 2 | 3 | 4 |

[0192] Two signals will be seen; the first signal corresponds to the ddATP filled in at position 1 complementary to the overhang and the second signal corresponds to the ddATP filled in at position 3 complementary to the overhang. The two signals can be separated based on molecular weight; allele 1 and allele 2 will be separated by two bases, which can be detected using any method that discriminates based on molecular weight.

[0193] Alternatively, if positions 2 and 3 are not complementary to adenine (*i.e.* positions 2 and 3 of the overhang correspond to guanine, cytosine, or adenine) but position 4 is complementary to adenine, labeled ddATP can be used to determine the sequence of both alleles.

[0194] Homozygous for adenine:

| | | | | | |
|-------------------|--------|----|---|---|---|
| Allele 1 | 5' CCC | A* | | | |
| | 3' GGG | T | G | G | T |
| Overhang position | | 1 | 2 | 3 | 4 |
| Allele 2 | 5' CCC | A* | | | |
| | 3' GGG | T | G | G | T |
| Overhang position | | 1 | 2 | 3 | 4 |

[0195] One signal will be seen that corresponds to the molecular weight of molecules filled in with ddATP at position one complementary to the overhang, which indicates that the individual is homozygous for adenine at the variable site.

[0196] Homozygous for guanine:

| | | | | | | |
|-------------------|--------|---|---|---|----|--|
| Allele 1 | 5' CCC | G | C | C | A* | |
| | 3' GGG | C | G | G | T | |
| Overhang position | | 1 | 2 | 3 | 4 | |
| Allele 2 | 5' CCC | G | C | C | A* | |
| | 3' GGG | C | G | G | T | |

| Overhang position | 1 | 2 | 3 | 4 |
|-------------------|---|---|---|---|
|-------------------|---|---|---|---|

[0197] One signal will be seen that corresponds to the molecular weight of molecules filled in at position 4 complementary to the overhang, which indicates that the individual is homozygous for guanine.

[0198] Heterozygous:

| | | | | | | |
|-------------------|--------|----|---|---|----|--|
| Allele 1 | 5' CCC | A* | | | | |
| | 3' GGG | T | G | G | T | |
| Overhang position | | 1 | 2 | 3 | 4 | |
| Allele 2 | 5' CCC | G | C | C | A* | |
| | 3' GGG | C | G | G | T | |
| Overhang position | | 1 | 2 | 3 | 4 | |

[0199] Two signals will be seen; the first signal corresponds to the ddATP filled in at position one complementary to the overhang and the second signal corresponds to the ddATP filled in at position 4 complementary to the overhang. The two signals can be separated based on molecular weight; allele 1 and allele 2 will be separated by three bases, which allows detection and quantitation of the signals. The molecules filled in at position 1 and those filled in at position 4 can be distinguished based on molecular weight.

[0200] As discussed above, if the variable site contains either adenine or guanine, either labeled adenine or labeled guanine can be used to determine the sequence of both alleles. If positions 2, 3, or 4 of the overhang are not complementary to adenine but one of the positions is complementary to a guanine, then labeled ddGTP can be used to determine whether the template DNA is homozygous or heterozygous for adenine or guanine. For example, if position 3 in the overhang corresponds to a cytosine then the following signals will be expected if the template DNA is homozygous for guanine, homozygous for adenine, or heterozygous:

[0201] Homozygous for guanine:

| | | | | | | |
|----------|--------|----|---|---|---|--|
| Allele 1 | 5' CCC | G* | | | | |
| | 3' GGG | C | T | C | T | |

| | | | | | |
|-------------------|--------|-----------|---|---|---|
| Overhang position | 1 | 2 | 3 | 4 | |
| Allele 2 | 5' CCC | G* | | | |
| | 3' GGG | C | T | C | T |
| Overhang position | 1 | 2 | 3 | 4 | |

[0202] One signal will be seen that corresponds to the molecular weight of molecules filled in with ddGTP at position one complementary to the overhang, which indicates that the individual is homozygous for guanine.

[0203] Homozygous for adenine:

| | | | | | |
|-------------------|--------|---|---|-----------|---|
| Allele 1 | 5' CCC | A | A | G* | |
| | 3' GGG | T | T | C | T |
| Overhang position | 1 | 2 | 3 | 4 | |
| Allele 2 | 5' CCC | A | A | G* | |
| | 3' GGG | T | T | C | T |
| Overhang position | 1 | 2 | 3 | 4 | |

[0204] One signal will be seen that corresponds to the molecular weight of molecules filled in at position 3 complementary to the overhang, which indicates that the individual is homozygous for adenine at the variable site.

[0205] Heterozygous:

| | | | | | |
|-------------------|--------|-----------|---|-----------|---|
| Allele 1 | 5' CCC | G* | | | |
| | 3' GGG | C | T | C | T |
| Overhang position | 1 | 2 | 3 | 4 | |
| Allele 2 | 5' CCC | A | A | G* | |
| | 3' GGG | T | T | C | T |
| Overhang position | 1 | 2 | 3 | 4 | |

[0206] Two signals will be seen; the first signal corresponds to the ddGTP filled in at position one complementary to the overhang and the second signal corresponds to the ddGTP filled in at position 3 complementary to the overhang. The two signals can be separated based on molecular weight; allele 1 and allele 2

will be separated by two bases, which allows easy detection and quantitation of the signals.

[0207] Some type IIS restriction enzymes also display alternative cutting as discussed above. For example, BsmFI will cut at 10/14 and 11/15 from the recognition site. However, the cutting patterns are not mutually exclusive; if the 11/15 cutting pattern is seen at a particular sequence, 10/14 cutting is also seen. If the restriction enzyme BsmF I cuts at 10/14 from the recognition site, the 5' overhang will be $X_1X_2X_3X_4$. If BsmF I cuts 11/15 from the recognition site, the 5' overhang will be $X_0X_1X_2X_3$. If position X_0 of the overhang is complementary to the labeled nucleotide, the labeled nucleotide will be incorporated at position X_0 and provides an additional level of quality assurance. It provides additional sequence information.

[0208] For example, if the variable site is adenine or guanine, and position 3 in the overhang is complementary to adenine, labeled ddATP can be used to determine the genotype at the variable site. If position 0 of the 11/15 overhang contains the nucleotide complementary to adenine, ddATP will be filled in and an additional signal will be seen.

[0209] Heterozygous:

| | | | | | | |
|-------------------|--------|-----------|---|-----------|---|--|
| 10/14 Allele 1 | 5' CCA | A* | | | | |
| | 3' GGT | T | G | T | G | |
| Overhang position | | 1 | 2 | 3 | 4 | |
| 10/14 Allele 2 | 5' CCA | G | C | A* | | |
| | 3' GGT | C | G | T | G | |
| Overhang position | | 1 | 2 | 3 | 4 | |
| 11/15 Allele 1 | 5' CC | A* | | | | |
| | 3' GG | T | T | G | T | |
| Overhang position | | 0 | 1 | 2 | 3 | |
| 11/15 Allele 2 | 5' CC | A* | | | | |
| | 3' GG | T | C | G | T | |

| Overhang position | 0 | 1 | 2 | 3 |
|-------------------|---|---|---|---|
|-------------------|---|---|---|---|

[0210] Three signals are seen; one corresponding to the ddATP incorporated at position 0 complementary to the overhang, one corresponding to the ddATP incorporated at position 1 complementary to the overhang, and one corresponding to the ddATP incorporated at position 3 complementary to the overhang. The molecules filled in at position 0, 1, and 3 complementary to the overhang differ in molecular weight and can be separated using any technique that discriminates based on molecular weight including but not limited to gel electrophoresis, and mass spectrometry.

[0211] For quantitating the ratio of one allele to another allele or when determining the relative amount of a mutant DNA sequence in the presence of wild type DNA sequence, an accurate and highly sensitive method of detection must be used. The alternate cutting displayed by type IIS restriction enzymes may increase the difficulty of determining ratios of one allele to another allele because the restriction enzyme may not display the alternate cutting (11/15) pattern on the two alleles equally. For example, allele 1 may be cut at 10/14 80% of the time, and 11/15 20% of the time. However, because the two alleles may differ in sequence, allele 2 may be cut at 10/14 90% of the time, and 11/15 20% of the time.

[0212] For purposes of quantitation, the alternate cutting problem can be eliminated when the nucleotide at position 0 of the overhang is not complementary to the labeled nucleotide. For example, if the variable site corresponds to adenine or guanine, and position 3 of the overhang is complementary to adenine (*i.e.*, a thymidine is located at position 3 of the overhang), labeled ddATP can be used to determine the genotype of the variable site. If position 0 of the overhang generated by the 11/15 cutting properties is not complementary to adenine, (*i.e.*, position 0 of the overhang corresponds to guanine, cytosine, or adenine) no additional signal will be seen from the fragments that were cut 11/15 from the recognition site. Position 0 complementary to the overhang can be filled in with unlabeled nucleotide,

eliminating any complexity seen from the alternate cutting pattern of restriction enzymes. This method provides a highly accurate method for quantitating the ratio of a variable site including but not limited to a mutation, or a single nucleotide polymorphism.

[0213] For instance, if SNP X can be adenine or guanine, this method of labeling allows quantitation of the alleles that correspond to adenine and the alleles that correspond to guanine, without determining if the restriction enzyme displays any differences between the alleles with regard to alternate cutting patterns.

[0214] Heterozygous:

| | | | | | | |
|-------------------|--------|-----------|----------|-----------|---|--|
| 10/14 Allele 1 | 5' CCG | A* | | | | |
| | 3' GGC | T | G | T | G | |
| Overhang position | | 1 | 2 | 3 | 4 | |
| 10/14 Allele 2 | 5' CCG | G | C | A* | | |
| | 3' GGC | C | G | T | G | |
| Overhang position | | 1 | 2 | 3 | 4 | |

[0215] The overhang generated by the alternate cutting properties of BsmF I is depicted below:

| | | | | | | |
|-------------------|-------|---|---|---|---|--|
| 11/15 Allele 1 | 5' CC | | | | | |
| | 3' GG | C | T | G | T | |
| Overhang position | | 0 | 1 | 2 | 3 | |

| | | | | | | |
|-------------------|-------|---|---|---|---|--|
| 11/15 Allele 2 | 5' CC | | | | | |
| | 3' GG | C | C | G | T | |
| Overhang position | | 0 | 1 | 2 | 3 | |

[0216] After filling in with labeled ddATP and unlabeled dGTP, dCTP, dTTP, the following molecules would be generated:

| | | | | |
|----------------|-------|----------|-----------|--|
| 11/15 Allele 1 | 5' CC | G | A* | |
|----------------|-------|----------|-----------|--|

| | | | | | |
|-------------------|-------|---|---|---|----|
| | 3' GG | C | T | G | T |
| Overhang position | | 0 | 1 | 2 | 3 |
| 11/15 Allele 2 | 5' CC | G | G | C | A* |
| | 3' GG | C | C | G | T |
| Overhang position | | 0 | 1 | 2 | 3 |

[0217] Two signals are seen; one corresponding to the molecules filled in with ddATP at position one complementary to the overhang and one corresponding to the molecules filled in with ddATP at position 3 complementary to the overhang. Position 0 of the 11/15 overhang is filled in with unlabeled nucleotide, which eliminates any difficulty in quantitating a ratio for the nucleotide at the variable site on allele 1 and the nucleotide at the variable site on allele 2.

[0218] Any nucleotide can be used including adenine, adenine derivatives, adenine homologues, guanine, guanine derivatives, guanine homologues, cytosine, cytosine derivatives, cytosine homologues, thymidine, thymidine derivatives, or thymidine homologues, or any combinations of adenine, adenine derivatives, adenine homologues, guanine, guanine derivatives, guanine homologues, cytosine, cytosine derivatives, cytosine homologues, thymidine, thymidine derivatives, or thymidine homologues.

[0219] The nucleotide can be labeled with any chemical group or moiety, including but not limited to radioactive molecules, fluorescent molecules, antibodies, antibody fragments, haptens, carbohydrates, biotin, derivatives of biotin, phosphorescent moieties, luminescent moieties, electrochemiluminescent moieties, chromatic moieties, and moieties having a detectable electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity. The nucleotide can be labeled with one or more than one type of chemical group or moiety.

[0220] In another embodiment, labeled and unlabeled nucleotides can be used. Any combination of deoxynucleotides and dideoxynucleotides can be used including but not limited to labeled dideoxynucleotides and labeled

deoxynucleotides; labeled dideoxynucleotides and unlabeled deoxynucleotides; unlabeled dideoxynucleotides and unlabeled deoxynucleotides; and unlabeled dideoxynucleotides and labeled deoxynucleotides.

[0221] In another embodiment, nucleotides labeled with a chemical moiety can be used in the PCR reaction. Unlabeled nucleotides then are used to fill-in the 5' overhangs generated after digestion with the restriction enzyme. An unlabeled terminating nucleotide can be used to in the presence of unlabeled nucleotides to determine the sequence of the alleles of a locus of interest.

[0222] For example, if labeled dTTP was used in the PCR reaction, the following 5' overhang would be generated after digestion with BsmF I:

| | | | | | | |
|-------------------|---------|---|---|---|---|--|
| 10/14 Allele 1 | 5' CT*G | A | | | | |
| | 3' GA C | T | G | T | G | |
| Overhang position | | 1 | 2 | 3 | 4 | |
| 10/14 Allele 2 | 5' CT*G | G | C | A | | |
| | 3' GA C | C | G | T | G | |
| Overhang position | | 1 | 2 | 3 | 4 | |

[0223] Unlabeled ddATP, unlabeled dCTP, unlabeled dGTP, and unlabeled dTTP can be used to fill-in the 5' overhang. Two signals will be generated; one signal corresponds to the DNA molecules filled in with unlabeled ddATP at position 1 complementary to the overhang and the second signal corresponds to DNA molecules filled in with unlabeled ddATP at position 3 complementary to the overhang. The DNA molecules can be separated based on molecular weight and can be detected by the fluorescence of the dTTP, which was incorporated during the PCR reaction.

[0224] The labeled DNA loci of interest sites can be analyzed by a variety of methods including but not limited to fluorescence detection, DNA sequencing gel, capillary electrophoresis on an automated DNA sequencing machine, microchannel electrophoresis, and other methods of sequencing, mass spectrometry, time of flight mass spectrometry, quadrupole mass spectrometry, magnetic sector mass spectrometry, electric sector mass spectrometry infrared

spectrometry, ultraviolet spectrometry, palentiostatic amperometry or by DNA hybridization techniques including Southern Blots, Slot Blots, Dot Blots, and DNA microarrays, wherein DNA fragments would be useful as both "probes" and "targets," ELISA, fluorimetry, and Fluorescence Resonance Energy Transfer (FRET).

[0225] This method of labeling is extremely sensitive and allows the detection of alleles of a locus of interest that are in various ratios including but not limited to 1:1, 1:2, 1:3, 1:4, 1:5, 1:6-1:10, 1:11-1:20, 1:21-1:30, 1:31-1:40, 1:41-1:50, 1:51-1:60, 1:61-1:70, 1:71-1:80, 1:81-1:90, 1:91:1:100, 1:101-1:200, 1:250, 1:251-1:300, 1:301-1:400, 1:401-1:500, 1:501-1:600, 1:601-1:700, 1:701-1:800, 1:801-1:900, 1:901-1:1000, 1:1001-1:2000, 1:2001-1:3000, 1:3001-1:4000, 1:4001-1:5000, 1:5001-1:6000, 1:6001-1:7000, 1:7001-1:8000, 1:8001-1:9000, 1:9001-1:10,000; 1:10,001-1:20,000, 1:20,001:1:30,000, 1:30,001-1:40,000, 1:40,001-1:50,000, and greater than 1:50,000.

[0226] For example, this method of labeling allows one nucleotide labeled with one signal generating moiety to be used to determine the sequence of alleles at a SNP locus, or detect a mutant allele amongst a population of normal alleles, or detect an allele encoding antibiotic resistance from a bacterial cell amongst alleles from antibiotic sensitive bacteria, or detect an allele from a drug resistant virus amongst alleles from drug-sensitive virus, or detect an allele from a non-pathogenic bacterial strain amongst alleles from a pathogenic bacterial strain.

[0227] As shown above, a single nucleotide can be used to determine the sequence of the alleles at a particular locus of interest. This method is especially useful for determining if an individual is homozygous or heterozygous for a particular mutation or to determine the sequence of the alleles at a particular SNP site. This method of labeling eliminates any errors caused by the quantum coefficients of various dyes. It also allows the reaction to proceed in a single reaction vessel including but not limited to a well of a microtiter plate, or a single eppendorf tube.

[0228] This method of labeling is especially useful for the detection of multiple genetic signals in the same sample. For example, this method is useful for the detection of fetal DNA in the blood, serum, or plasma of a pregnant female, which contains both maternal DNA and fetal DNA. The maternal DNA and fetal DNA may be present in the blood, serum or plasma at ratios such as 97:3; however, the above-described method can be used to detect the fetal DNA. This method of labeling can be used to detect two, three, or four different genetic signals in the sample population

[0229] This method of labeling is especially useful for the detection of a mutant allele that is among a large population of wild type alleles. Furthermore, this method of labeling allows the detection of a single mutant cell in a large population of wild type cells. For example, this method of labeling can be used to detect a single cancerous cell among a large population of normal cells. Typically, cancerous cells have mutations in the DNA sequence. The mutant DNA sequence can be identified even if there is a large background of wild type DNA sequence. This method of labeling can be used to screen, detect, or diagnosis any type of cancer including but not limited to colon, renal, breast, bladder, liver, kidney, brain, lung, prostate, and cancers of the blood including leukemia.

[0230] This labeling method can also be used to detect pathogenic organisms, including but not limited to bacteria, fungi, viruses, protozoa, and mycobacteria. It can also be used to discriminate between pathogenic strains of microorganism and non-pathogenic strains of microorganisms including but not limited to bacteria, fungi, viruses, protozoa, and mycobacteria.

[0231] For example, there are several strains of *Escherichia coli* (*E. coli*), and most are non-pathogenic. However, several strains, such as *E. coli* O157 are pathogenic. There are genetic differences between non-pathogenic *E. coli* strains and pathogenic *E. coli*. The above described method of labeling can be used to detect pathogenic microorganisms in a large population of non-pathogenic organisms, which are sometimes associated with the normal flora of an individual.

[0232] In another embodiment, the sequence of the locus of interest can be determined by detecting the incorporation of a nucleotide that is 3' to the locus of interest, wherein said nucleotide is a different nucleotide from the possible nucleotides at the locus of interest. This embodiment is especially useful for the sequencing and detection of SNPs. The efficiency and rate at which DNA polymerases incorporate nucleotides varies for each nucleotide.

[0233] According to the data from the Human Genome Project, 99% of all SNPs are binary. The sequence of the human genome can be used to determine the nucleotide that is 3' to the SNP of interest. When the nucleotide that is 3' to the SNP site differs from the possible nucleotides at the SNP site, a nucleotide that is one or more than one base 3' to the SNP can be used to determine the identity of the SNP.

[0234] For example, suppose the identity of SNP X on chromosome 13 is to be determined. The sequence of the human genome indicates that SNP X can either be adenine or guanine and that a nucleotide 3' to the locus of interest is a thymidine. A primer that contains a restriction enzyme recognition site for BsmF I, which is designed to be 13 bases from the locus of interest after amplification, is used to amplify a DNA fragment containing SNP X. Digestion with the restriction enzyme BsmF I generates a 5' overhang that contains the locus of interest, which can either be adenine or guanine. The digestion products can be split into two "fill in" reactions: one contains dTTP, and the other reaction contains dCTP. If the locus of interest is homozygous for guanine, only the DNA molecules that were mixed with dCTP will be filled in. If the locus of interest is homozygous for adenine, only the DNA molecules that were mixed with dTTP will be filled in. If the locus of interest is heterozygous, the DNA molecules that were mixed with dCTP will be filled in as well as the DNA molecules that were mixed with dTTP. After washing to remove the excess dNTP, the samples are filled in with labeled ddATP, which is complementary to the nucleotide (thymidine) that is 3' to the locus of interest. The DNA molecules that were filled in by the previous reaction will be filled in with labeled ddATP. If the individual

is homozygous for adenosine, the DNA molecules that were mixed with dTTP subsequently will be filled in with the labeled ddATP. However, the DNA molecules that were mixed with dCTP, would not have incorporated that nucleotide, and therefore, could not incorporate the ddATP. Detection of labeled ddATP only in the molecules that were mixed with dTTP indicates that the identity of the nucleotide at SNP X on chromosome 13 is adenosine.

[0235] In another embodiment, large scale screening for the presence or absence of single nucleotide mutations can be performed. One to tens to hundreds to thousands of loci of interest on a single chromosome or on multiple chromosomes can be amplified with primers as described above in the "Primer Design" section. The primers can be designed so that each amplified loci of interest is of a different size (FIG. 2). The amplified loci of interest that are predicted, based on the published wild type sequences, to have the same nucleotide at the locus of interest can be pooled together, bound to a solid support, including wells of a microtiter plate coated with streptavidin, and digested with the restriction enzyme that will bind the recognition site on the second primer. After digestion, the 3' recessed end can be filled in with a mixture of labeled ddATP, ddTTP, ddGTP, ddCTP, where each nucleotide is labeled with a different group. After washing to remove the excess nucleotide, the fluorescence spectra can be detected using a plate reader or fluorimeter directly on the streptavidin coated plates. If all 50 loci of interest contain the wild type nucleotide, only one fluorescence spectra will be seen. However, if one or more than one of the 50 loci of interest contain a mutation, a different nucleotide will be incorporated and other fluorescence pattern(s) will be seen. The nucleotides can be released from the solid matrix, and analyzed on a sequencing gel to determine the loci of interest that contained the mutations. As each of the 50 loci of interest are of different size, they will separate on a sequencing gel.

[0236] The multiple loci of interest can be of a DNA sample from one individual representing multiple loci of interest on a single chromosome, multiple chromosomes, multiple genes, a single gene, or any combination thereof. The

multiple loci of interest also can represent the same locus of interest but from multiple individuals. For example, 50 DNA samples from 50 different individuals can be pooled and analyzed to determine a particular nucleotide of interest at gene "X."

[0237] When human data is being analyzed, the known sequence can be a specific sequence that has been determined from one individual (including e.g. the individual whose DNA is currently being analyzed), or it can be a consensus sequence such as that published as part of the human genome.

Kits

[0238] The methods of the invention are most conveniently practiced by providing the reagents used in the methods in the form of kits. A kit preferably contains one or more of the following components: written instructions for the use of the kit, appropriate buffers, salts, DNA extraction detergents, primers, nucleotides, labeled nucleotides, 5' end modification materials, and if desired, water of the appropriate purity, confined in separate containers or packages, such components allowing the user of the kit to extract the appropriate nucleic acid sample, and analyze the same according to the methods of the invention. The primers that are provided with the kit will vary, depending upon the purpose of the kit and the DNA that is desired to be tested using the kit. In preferred embodiments the kits contain a primer that allows the generation of a recognition site for a restriction enzyme such that digestion with the enzyme generates in the DNA fragment generated during the sequencing method, a 5' overhang containing the locus of interest.

[0239] A kit can also be designed to detect a desired or variety of single nucleotide polymorphisms, especially those associated with an undesired condition or disease. For example, one kit can comprise, among other components, a set or sets of primers to amplify one or more loci of interest associated with breast cancer. Another kit can comprise, among other components, a set or sets of primers for genes associated with a predisposition to develop type I or type II diabetes. Still, another kit can comprise, among other

components, a set or sets of primers for genes associated with a predisposition to develop heart disease. Details of utilities for such kits are provided in the "Utilities" section below.

Utilities

[0240] The methods of the invention can be used whenever it is desired to know the sequence of a certain nucleic acid, locus of interest or loci of interest therein. The method of the invention is especially useful when applied to genomic DNA. When DNA from an organism-specific or species-specific locus or loci of interest is amplified, the method of the invention can be used in genotyping for identification of the source of the DNA, and thus confirm or provide the identity of the organism or species from which the DNA sample was derived. The organism can be any nucleic acid containing organism, for example, virus, bacterium, yeast, plant, animal or human.

[0241] Within any population of organisms, the method of the invention is useful to identify differences between the sequence of the sample nucleic acid and that of a known nucleic acid. Such differences can include, for example, allelic variations, mutations, polymorphisms and especially single nucleotide polymorphisms.

[0242] In a preferred embodiment, the method of the invention provides a method for identification of single nucleotide polymorphisms.

[0243] In a preferred embodiment, the method of the invention provides a method for identification of the presence of a disease, especially a genetic disease that arises as a result of the presence of a genomic sequence, or other biological condition that it is desired to identify in an individual for which it is desired to know the same. The identification of such sequence in the subject based on the presence of such genomic sequence can be used, for example, to determine if the subject is a carrier or to assess if the subject is predisposed to developing a certain genetic trait, condition or disease. The method of the invention is especially useful in prenatal genetic testing of parents and child. Examples of some of the

diseases that can be diagnosed by this invention are listed in Table II.

TABLE II

| |
|---|
| Achondroplasia |
| Adrenoleukodystrophy, X-Linked |
| Agammaglobulinemia, X-Linked |
| Alagille Syndrome |
| Alpha-Thalassemia X-Linked Mental Retardation Syndrome |
| Alzheimer Disease |
| Alzheimer Disease, Early-Onset Familial |
| Amyotrophic Lateral Sclerosis Overview |
| Androgen Insensitivity Syndrome |
| Angelman Syndrome |
| Ataxia Overview, Hereditary |
| Ataxia-Telangiectasia |
| Becker Muscular Dystrophy (also The Dystrophinopathies) |
| Beckwith-Wiedemann Syndrome |
| Beta-Thalassemia |
| Biotinidase Deficiency |
| Branchiootorenal Syndrome |
| BRCA1 and BRCA2 Hereditary Breast/Ovarian Cancer |
| Breast Cancer |
| CADASIL |
| Canavan Disease |
| Cancer |
| Charcot-Marie-Tooth Hereditary Neuropathy |
| Charcot-Marie-Tooth Neuropathy Type 1 |
| Charcot-Marie-Tooth Neuropathy Type 2 |
| Charcot-Marie-Tooth Neuropathy Type 4 |
| Charcot-Marie-Tooth Neuropathy Type X |
| Cockayne Syndrome |
| Colon Cancer |
| Contractural Arachnodactyly, Congenital |
| Craniosynostosis Syndromes (FGFR-Related) |
| Cystic Fibrosis |
| Cystinosis |
| Deafness and Hereditary Hearing Loss |
| DRPLA (Dentatorubral-Pallidoluysian Atrophy) |
| DiGeorge Syndrome (also 22q11 Deletion Syndrome) |
| Dilated Cardiomyopathy, X-Linked |
| Down Syndrome (Trisomy 21) |
| Duchenne Muscular Dystrophy (also The Dystrophinopathies) |
| Dystonia, Early-Onset Primary (DYT1) |
| Dystrophinopathies, The |
| Ehlers-Danlos Syndrome, Kyphoscoliotic Form |

| |
|--|
| Ehlers-Danlos Syndrome, Vascular Type |
| Epidermolysis Bullosa Simplex |
| Exostoses, Hereditary Multiple |
| Faciocapulohumeral Muscular Dystrophy |
| Factor V Leiden Thrombophilia |
| Familial Adenomatous Polyposis (FAP) |
| Familial Mediterranean Fever |
| Fragile X Syndrome |
| Friedreich Ataxia |
| Frontotemporal Dementia with Parkinsonism-17 |
| Galactosemia |
| Gaucher Disease |
| Hemochromatosis, Hereditary |
| Hemophilia A |
| Hemophilia B |
| Hemorrhagic Telangiectasia, Hereditary |
| Hearing Loss and Deafness, Nonsyndromic, DFNA3 (Connexin 26) |
| Hearing Loss and Deafness, Nonsyndromic, DFNB1 (Connexin 26) |
| Hereditary Spastic Paraparesis |
| Hermansky-Pudlak Syndrome |
| Hexosaminidase A Deficiency (also Tay-Sachs) |
| Huntington Disease |
| Hypochondroplasia |
| Ichthyosis, Congenital, Autosomal Recessive |
| Incontinentia Pigmenti |
| Kennedy Disease (also Spinal and Bulbar Muscular Atrophy) |
| Krabbe Disease |
| Leber Hereditary Optic Neuropathy |
| Lesch-Nyhan Syndrome |
| Leukemias |
| Li-Fraumeni Syndrome |
| Limb-Girdle Muscular Dystrophy |
| Lipoprotein Lipase Deficiency, Familial |
| Lissencephaly |
| Marfan Syndrome |
| MELAS (Mitochondrial Encephalomyopathy, Lactic Acidosis, and Stroke-Like Episodes) |
| Monosomies |
| Multiple Endocrine Neoplasia Type 2 |
| Multiple Exostoses, Hereditary |
| Muscular Dystrophy, Congenital |
| Myotonic Dystrophy |
| Nephrogenic Diabetes Insipidus |
| Neurofibromatosis 1 |
| Neurofibromatosis 2 |
| Neuropathy with Liability to Pressure Palsies, Hereditary |
| Niemann-Pick Disease Type C |

| |
|--|
| Nijmegen Breakage Syndrome |
| Norrie Disease |
| Oculocutaneous Albinism Type 1 |
| Oculopharyngeal Muscular Dystrophy |
| Ovarian Cancer |
| Pallister-Hall Syndrome |
| Parkin Type of Juvenile Parkinson Disease |
| Pelizaeus-Merzbacher Disease |
| Pendred Syndrome |
| Peutz-Jeghers Syndrome |
| Phenylalanine Hydroxylase Deficiency |
| Prader-Willi Syndrome |
| <i>PROP1</i> -Related Combined Pituitary Hormone Deficiency (CPHD) |
| Prostate Cancer |
| Retinitis Pigmentosa |
| Retinoblastoma |
| Rothmund-Thomson Syndrome |
| Smith-Lemli-Opitz Syndrome |
| Spastic Paraplegia, Hereditary |
| Spinal and Bulbar Muscular Atrophy (also Kennedy Disease) |
| Spinal Muscular Atrophy |
| Spinocerebellar Ataxia Type 1 |
| Spinocerebellar Ataxia Type 2 |
| Spinocerebellar Ataxia Type 3 |
| Spinocerebellar Ataxia Type 6 |
| Spinocerebellar Ataxia Type 7 |
| Stickler Syndrome (Hereditary Arthroophthalmopathy) |
| Tay-Sachs (also GM2 Gangliosidoses) |
| Trisomies |
| Tuberous Sclerosis Complex |
| Usher Syndrome Type I |
| Usher Syndrome Type II |
| Velocardiofacial Syndrome (also 22q11 Deletion Syndrome) |
| Von Hippel-Lindau Syndrome |
| Williams Syndrome |
| Wilson Disease |
| X-Linked Adrenoleukodystrophy |
| X-Linked Agammaglobulinemia |
| X-Linked Dilated Cardiomyopathy (also The Dystrophinopathies) |
| X-Linked Hypotonic Facies Mental Retardation Syndrome |

[0244] The method of the invention is useful for screening an individual at multiple loci of interest, such as tens, hundreds, or even thousands of loci of interest associated with a genetic trait or genetic disease by sequencing the loci of interest that are associated with the trait or disease state, especially those most

frequently associated with such trait or condition. The invention is useful for analyzing a particular set of diseases including but not limited to heart disease, cancer, endocrine disorders, immune disorders, neurological disorders, musculoskeletal disorders, ophthalmologic disorders, genetic abnormalities, trisomies, monosomies, transversions, translocations, skin disorders, and familial diseases.

[0245] The method of the invention can be used to genotype microorganisms so as to rapidly identify the presence of a specific microorganism in a substance, for example, a food substance. In that regard, the method of the invention provides a rapid way to analyze food, liquids or air samples for the presence of an undesired biological contamination, for example, microbiological, fungal or animal waste material. The invention is useful for detecting a variety of organisms, including but not limited to bacteria, viruses, fungi, protozoa, molds, yeasts, plants, animals, and archaebacteria. The invention is useful for detecting organisms collected from a variety of sources including but not limited to water, air, hotels, conference rooms, swimming pools, bathrooms, aircraft, spacecraft, trains, buses, cars, offices, homes, businesses, churches, parks, beaches, athletic facilities, amusement parks, theaters, and any other facility that is a meeting place for the public.

[0246] The method of the invention can be used to test for the presence of many types of bacteria or viruses in blood cultures from human or animal blood samples.

[0247] The method of the invention can also be used to confirm or identify the presence of a desired or undesired yeast strain, or certain traits thereof, in fermentation products, e.g. wine, beer, and other alcohols or to identify the absence thereof.

[0248] The method of the invention can also be used to confirm or identify the relationship of a DNA of unknown sequence to a DNA of known origin or sequence, for example, for use in criminology, forensic science, maternity or paternity testing, archeological analysis, and the like.

[0249] The method the invention can also be used to determine the genotypes of plants, trees and bushes, and hybrid plants, trees and bushes, including plants, trees and bushes that produce fruits and vegetables and other crops, including but not limited to wheat, barley, corn, tobacco, alfalfa, apples, apricots, bananas, oranges, pears, nectarines, figs, dates, raisins, plums, peaches, apricots, blueberries, strawberries, cranberries, berries, cherries, kiwis, limes, lemons, melons, pineapples, plantains, guavas, prunes, passion fruit, tangerines, grapefruit, grapes, watermelon, cantaloupe, honeydew melons, pomegranates, persimmons, nuts, artichokes, bean sprouts, beets, cardoon, chayote, endive, leeks, okra, green onions, scallions, shallots, parsnips, sweet potatoes, yams, asparagus, avocados, kohlrabi, rutabaga, eggplant, squash, turnips, pumpkins, tomatoes, potatoes, cucumbers, carrots, cabbage, celery, broccoli, cauliflower, radishes, peppers, spinach, mushrooms, zucchini, onions, peas, beans, and other legumes.

[0250] Especially, the method of the invention is useful to screen a mixture of nucleic acid samples that contain many different loci of interest and/or a mixture of nucleic acid samples from different sources that are to be analyzed for a locus of interest. Examples of large scale screening include taking samples of nucleic acid from herds of farm animals, or crops of food plants such as, for example, corn or wheat, pooling the same, and then later analyzing the pooled samples for the presence of an undesired genetic marker, with individual samples only being analyzed at a later date if the pooled sample indicates the presence of such undesired genetic sequence. An example of an undesired genetic sequence would be the detection of viral or bacterial nucleic acid sequence in the nucleic acid samples taken from the farm animals, for example, mycobacterium or hoof and mouth disease virus sequences or fungal or bacterial pathogen of plants.

[0251] Another example where pools of nucleic acid can be used is to test for the presence of a pathogen or gene mutation in samples from one or more tissues from an animal or human subject, living or dead, especially a subject who can be in need of treatment if the pathogen or mutation is detected. For example, numerous samples can be taken from an animal or human subject to be screened

for the presence of a pathogen or otherwise undesired genetic mutation, the loci of interest from each biological sample amplified individually, and then samples of the amplified DNA combined for the restriction digestion, "filling in," and detection. This would be useful as an initial screening for the assay of the presence or absence of nucleic acid sequences that would be diagnostic of the presence of a pathogen or mutation. Then, if the undesired nucleic acid sequence of the pathogen or mutation was detected, the individual samples could be separately analyzed to determine the distribution of the undesired sequence. Such an analysis is especially cost effective when there are large numbers of samples to be assayed. Samples of pathogens include the mycobacteria, especially those that cause tuberculosis or paratuberculosis, bacteria, especially bacterial pathogens used in biological warfare, including *Bacillus anthracis*, and virulent bacteria capable of causing food poisoning, viruses, especially the influenza and AIDS virus, and mutations known to be associated with malignant cells. Such an analysis would also be advantageous for the large scale screening of food products for pathogenic bacteria.

[0252] Conversely, the method of the invention can be used to detect the presence and distribution of a desired genetic sequence at various locations in a plant, animal or human subject, or in a population of subjects, e.g. by screening of a combined sample followed by screening of individual samples, as necessary.

[0253] The method of the invention is useful for analyzing genetic variations of an individual that have an effect on drug metabolism, drug interactions, and the responsiveness to a drug or to multiple drugs. The method of the invention is especially useful in pharmacogenomics.

[0254] Having now generally described the invention, the same will become better understood by reference to certain specific examples which are included herein for purposes of illustration only and are not intended to be limiting unless otherwise specified.

EXAMPLES

[0255] The following examples are illustrative only and are not intended to limit the scope of the invention as defined by the claims.

EXAMPLE 1

[0256] DNA sequences were amplified by PCR, wherein the annealing step in cycle 1 was performed at a specified temperature, and then increased in cycle 2, and further increased in cycle 3 for the purpose of reducing non-specific amplification. The TM1 of cycle 1 of PCR was determined by calculating the melting temperature of the 3' region, which anneals to the template DNA, of the second primer. For example, in FIG. 1B, the TM1 can be about the melting temperature of region "c." The annealing temperature was raised in cycle 2, to TM2, which was about the melting temperature of the 3' region, which anneals to the template DNA, of the first primer. For example, in FIG. 1C, the annealing temperature (TM2) corresponds to the melting temperature of region "b'". In cycle 3, the annealing temperature was raised to TM3, which was about the melting temperature of the entire sequence of the second primer. For example, in FIG. 1D, the annealing temperature (TM3) corresponds to the melting temperature of region "c" + region "d". The remaining cycles of amplification were performed at TM3.

Preparation of Template DNA

[0257] The template DNA was prepared from a 5 ml sample of blood obtained by venipuncture from a human volunteer with informed consent. Blood was collected from 36 volunteers. Template DNA was isolated from each blood sample using QIAamp DNA Blood Midi Kit supplied by QIAGEN (Catalog number 51183). Following isolation, the template DNA from each of the 36 volunteers was pooled for further analysis.

Design of Primers

[0258] The following four single nucleotide polymorphisms were analyzed: SNP HC21S00340, identification number as assigned by Human Chromosome 21

cSNP Database, (FIG. 3, lane 1) located on chromosome 21; SNP TSC 0095512 (FIG. 3, lane 2) located on chromosome 1, SNP TSC 0214366 (FIG. 3, lane 3) located on chromosome 1; and SNP TSC 0087315 (FIG. 3, lane 4) located on chromosome 1. The SNP Consortium Ltd database can be accessed at <http://snp.cshl.org/>, website address effective as of February 14, 2002.

[0259] SNP HC21S00340 was amplified using the following primers:

First primer:

5' TAGAATAGCACTGAATTCAAGGAATAACATTGTCAC 3' (SEQ ID NO:9)

Second primer:

5' ATCACGATAAACGGCCAAACTCAGGTTA 3' (SEQ ID NO:10)

[0260] SNP TSC0095512 was amplified using the following primers:

First primer:

5' AAGTTTAGATCAGAATTCTGTGAAAGCAGAAGTTGTCTG 3' (SEQ ID NO:11)

Second primer:

5' TCTCCAACTAACGGCTCATCGAGTAAAG 3' (SEQ ID NO:12)

[0261] SNP TSC0214366 was amplified using the following primers:

First primer:

5' ATGACTAGCTATGAATTCTGTTCAAGGTAGAAAATGGAA 3' (SEQ ID NO:13)

Second primer:

5' GAGAATTAGAACGGCCAAATCCCACTC 3' (SEQ ID NO:14)

[0262] SNP TSC 0087315 was amplified using the following primers:

First primer:

5' TTACAATGCATGAATTCTGTTCTCAAAGTGC 3' (SEQ ID NO:15)

Second primer:

5' TGGACCATAAACGGCCAAAAACTGTAAG 3' (SEQ ID NO:16)

[0263] All primers were designed such that the 3' region was complementary to either the upstream or downstream sequence flanking each locus of interest and the 5' region contained a restriction enzyme recognition site. The first primer contained a biotin tag at the 5' end and a recognition site for the restriction enzyme EcoRI. The second primer contained the recognition site for the restriction enzyme BceA I.

PCR Reaction

[0264] All four loci of interest were amplified from the template genomic DNA using PCR (U.S. Patent Nos. 4,683,195 and 4,683,202). The components of the PCR reaction were as follows: 40 ng of template DNA, 5 μ M first primer, 5 μ M second primer, 1X HotStarTaq Master Mix as obtained from QIAGEN (Catalog No. 203443). The HotStarTaq Master Mix contained DNA polymerase, PCR buffer, 200 μ M of each dNTP, and 1.5 mM MgCl₂.

[0265] Amplification of each template DNA that contained the SNP of interest was performed using three different series of annealing temperatures, herein referred to as low stringency annealing temperature, medium stringency annealing temperature, and high stringency annealing temperature. Regardless of the annealing temperature protocol, each PCR reaction consisted of 40 cycles of amplification. PCR reactions were performed using the HotStarTaq Master Mix Kit supplied by QIAGEN. As instructed by the manufacturer, the reactions were incubated at 95°C for 15 min. prior to the first cycle of PCR. The denaturation step after each extension step was performed at 95°C for 30 sec. The annealing reaction was performed at a temperature that permitted efficient extension without any increase in temperature.

[0266] The low stringency annealing reaction comprised three different annealing temperatures in each of the first three cycles. The annealing temperature for the first cycle was 37°C for 30 sec.; the annealing temperature for the second cycle was 57°C for 30 sec.; the annealing temperature for the third cycle was 64°C for 30 sec. Annealing was performed at 64°C for subsequent cycles until completion.

[0267] As shown in the photograph of the gel (FIG. 3A), multiple bands were observed after amplification of the DNA template containing SNP TSC 0087315 (lane 4). Amplification of the DNA templates containing SNP HC21S00340 (lane 1), SNP TSC0095512 (lane 2), and SNP TSC0214366 (lane 3) generated a single band of high intensity and one band of faint intensity, which was of higher molecular weight. When the low annealing temperature conditions were used, the correct size product was generated and this was the predominant product in each reaction.

[0268] The medium stringency annealing reaction comprised three different annealing temperatures in each of the first three cycles. The annealing temperature for the first cycle was 40°C for 30 seconds; the annealing temperature for the second cycle was 60°C for 30 seconds; and the annealing temperature for the third cycle was 67°C for 30 seconds. Annealing was performed at 67°C for subsequent cycles until completion. Similar to what was observed under low stringency annealing conditions, amplification of the DNA template containing SNP TSC0087315 (FIG. 3B, lane 4) generated multiple bands under conditions of medium stringency. Amplification of the other three DNA fragments containing SNPs (lanes 1-3) produced a single band. These results demonstrate that variable annealing temperatures can be used to cleanly amplify loci of interest from genomic DNA with a primer that has an annealing length of 13 bases.

[0269] The high stringency annealing reaction was comprised of three different annealing temperatures in each of the first three cycles. The annealing temperature of the first cycle was 46°C for 30 seconds; the annealing temperature of the second cycle was 65°C for 30 seconds; and the annealing temperature for the third cycle was 72°C for 30 seconds. Annealing was performed at 72°C for subsequent cycles until completion. As shown in the photograph of the gel (FIG. 3C), amplification of the DNA template containing SNP TSC0087315 (lane 4) using the high stringency annealing temperatures generated a single band of the correct molecular weight. By raising the annealing temperatures for each of the first three cycles, non-specific amplification was eliminated. Amplification of the

DNA fragment containing SNP TSC0095512 (lane 2) generated a single band. DNA fragments containing SNPs HC21S00340 (lane 1), and TSC0214366 (lane 3) failed to amplify at the high stringency annealing temperatures, however, at the medium stringency annealing temperatures, these DNA fragments containing SNPs amplified as a single band. These results demonstrate that variable annealing temperatures can be used to reduce non-specific PCR products, as demonstrated for the DNA fragment containing SNP TSC0087315 (FIG. 3, lane 4).

EXAMPLE 2

[0270] SNPs on chromosomes 1 (TSC0095512), 13 (TSC0264580), and 21 (HC21S00027) were analyzed. SNP TSC0095512 was analyzed using two different sets of primers, and SNP HC21S00027 was analyzed using two types of reactions for the incorporation of nucleotides.

Preparation of Template DNA

[0271] The template DNA was prepared from a 5 ml sample of blood obtained by venipuncture from a human volunteer with informed consent. Template DNA was isolated using the QIAamp DNA Blood Midi Kit supplied by QIAGEN (Catalog number 51183). The template DNA was isolated as per instructions included in the kit. Following isolation, template DNA from thirty-six human volunteers were pooled together and cut with the restriction enzyme EcoRI. The restriction enzyme digestion was performed as per manufacturer's instructions.

Design of Primers

[0272] SNP HC21S00027 was amplified by PCR using the following primer set:

First primer:

5' ATAACCGTATGCGAATTCTATAATTTCTGATAAAGG 3' (SEQ ID NO:17)

Second primer:

5' CTTAAATCAGGGACTAGGTAAACTTCA 3' (SEQ ID NO:18)

[0273] The first primer contained a biotin tag at the extreme 5' end, and the nucleotide sequence for the restriction enzyme EcoRI. The second primer contained the nucleotide sequence for the restriction enzyme BsmF I (FIG. 4A).

[0274] Also, SNP HC21S00027 was amplified by PCR using the same first primer but a different second primer with the following sequence:

Second primer:

5' CTTAAATCAGACGGCTAGGTAAACTTCA 3' (SEQ ID NO:19)

[0275] This second primer contained the recognition site for the restriction enzyme BceA I (FIG. 4B).

SNP TSC0095512 was amplified by PCR using the following primers:

First primer:

5' AAGTTTAGATCAGAATTCTGTGAAAGCAGAAGTTGTCTG 3' (SEQ ID NO:11)

Second primer:

5' TCTCCAACTAGGGACTCATCGAGTAAAG 3' (SEQ ID NO:20)

[0276] The first primer had a biotin tag at the 5' end and contained a restriction enzyme recognition site for EcoRI. The second primer contained a restriction enzyme recognition site for BsmF I (FIG. 4C).

[0277] Also, SNP TSC0095512 was amplified using the same first primer and a different second primer with the following sequence:

Second primer:

5' TCTCCAACTAACGGCTCATCGAGTAAAG 3' (SEQ ID NO: 12)

[0278] This second primer contained the recognition site for the restriction enzyme BceA I (FIG. 4D).

[0279] SNP TSC0264580, which is located on chromosome 13, was amplified with the following primers:

First primer:

5' AACGCCGGGCGAGAATTCACTTGCAAGG 3' (SEQ ID NO:21)

Second primer:

5' CTACACATATCTGGGACGTTGGCCATCC 3' (SEQ ID NO:22)

[0280] The first primer contained a biotin tag at the extreme 5' end and had a restriction enzyme recognition site for EcoRI. The second primer contained a restriction enzyme recognition site for BsmF I.

PCR Reaction

[0281] All loci of interest were amplified from the template genomic DNA using the polymerase chain reaction (PCR, U.S. Patent Nos. 4,683,195 and 4,683,202, incorporated herein by reference). In this example, the loci of interest were amplified in separate reaction tubes but they could also be amplified together in a single PCR reaction. For increased specificity, a "hot-start" PCR was used. PCR reactions were performed using the HotStarTaq Master Mix Kit supplied by QIAGEN (catalog number 203443). The amount of template DNA and primer per reaction can be optimized for each locus of interest but in this example, 40 ng of template human genomic DNA and 5 μ M of each primer were used. Forty cycles of PCR were performed. The following PCR conditions were used:

- (1) 95°C for 15 minutes and 15 seconds;
- (2) 37°C for 30 seconds;
- (3) 95°C for 30 seconds;
- (4) 57°C for 30 seconds;
- (5) 95°C for 30 seconds;
- (6) 64°C for 30 seconds;
- (7) 95°C for 30 seconds;
- (8) Repeat steps 6 and 7 thirty nine (39) times;
- (9) 72°C for 5 minutes.

[0282] In the first cycle of PCR, the annealing temperature was about the melting temperature of the 3' annealing region of the second primers, which was 37°C. The annealing temperature in the second cycle of PCR was about the melting temperature of the 3' region, which anneals to the template DNA, of the first primer, which was 57°C. The annealing temperature in the third cycle of

PCR was about the melting temperature of the entire sequence of the second primer, which was 64°C. The annealing temperature for the remaining cycles was 64°C. Escalating the annealing temperature from TM1 to TM2 to TM3 in the first three cycles of PCR greatly improves specificity. These annealing temperatures are representative, and the skilled artisan will understand the annealing temperatures for each cycle are dependent on the specific primers used.

[0283] The temperatures and times for denaturing, annealing, and extension, can be optimized by trying various settings and using the parameters that yield the best results. Schematics of the PCR products for SNP HC21S00027 and SNP TSC095512 are shown in FIGS. 5A-5D.

Purification of Fragment Containing Locus of Interest

[0284] The PCR products were separated from the genomic template DNA. Each PCR product was divided into four separate reaction wells of a Streptawell, transparent, High-Bind plate from Roche Diagnostics GmbH (catalog number 1 645 692, as listed in Roche Molecular Biochemicals, 2001 Biochemicals Catalog). The first primers contained a 5' biotin tag so the PCR products bound to the Streptavidin coated wells while the genomic template DNA did not. The streptavidin binding reaction was performed using a Thermomixer (Eppendorf) at 1000 rpm for 20 min. at 37°C. Each well was aspirated to remove unbound material, and washed three times with 1X PBS, with gentle mixing (Kandpal et al., Nucl. Acids Res. 18:1789-1795 (1990); Kaneoka et al., Biotechniques 10:30-34 (1991); Green et al., Nucl. Acids Res. 18:6163-6164 (1990)).

Restriction Enzyme Digestion of Isolated Fragments Containing Loci of Interest

[0285] The purified PCR products were digested with the restriction enzyme that bound the recognition site incorporated into the PCR products from the second primer. DNA templates containing SNP HC21S00027 (FIG. 6A and 6B) and SNP TSC0095512 (FIG. 6C and 6D) were amplified in separate reactions using two different second primers. FIG. 6A (SNP HC21S00027) and FIG. 6C (SNP TSC0095512) depict the PCR products after digestion with the restriction

enzyme BsmF I (New England Biolabs catalog number R0572S). FIG. 6B (SNP HC21S00027) and FIG. 6D (SNP TSC0095512) depict the PCR products after digestion with the restriction enzyme BceA I (New England Biolabs, catalog number R0623S). The digests were performed in the Streptawells following the instructions supplied with the restriction enzyme. The DNA fragment containing SNP TSC0264580 was digested with BsmF I. After digestion with the appropriate restriction enzyme, the wells were washed three times with PBS to remove the cleaved fragments.

Incorporation of Labeled Nucleotide

[0286] The restriction enzyme digest described above yielded a DNA fragment with a 5' overhang, which contained the SNP site or locus of interest and a 3' recessed end. The 5' overhang functioned as a template allowing incorporation of a nucleotide or nucleotides in the presence of a DNA polymerase.

[0287] For each SNP, four separate fill in reactions were performed; each of the four reactions contained a different fluorescently labeled ddNTP (ddATP, ddTTP, ddGTP, or ddCTP). The following components were added to each fill in reaction: 1 μ l of a fluorescently labeled ddNTP, 0.5 μ l of unlabeled ddNTPs (40 μ M), which contained all nucleotides except the nucleotide that was fluorescently labeled, 2 μ l of 10X sequenase buffer, 0.25 μ l of Sequenase, and water as needed for a 20 μ l reaction. All of the fill in reactions were performed at 40°C for 10 min. Non-fluorescently labeled ddNTP was purchased from Fermentas Inc. (Hanover, MD). All other labeling reagents were obtained from Amersham (Thermo Sequenase Dye Terminator Cycle Sequencing Core Kit, US 79565). In the presence of fluorescently labeled ddNTPs, the 3' recessed end was extended by one base, which corresponds to the SNP or locus of interest (FIG 7A-7D).

[0288] A mixture of labeled ddNTPs and unlabeled dNTPs also was used for the "fill in" reaction for SNP HC21S00027. The "fill in" conditions were as described above except that a mixture containing 40 μ M unlabeled dNTPs, 1 μ l fluorescently labeled ddATP, 1 μ l fluorescently labeled ddTTP, 1 μ l fluorescently labeled ddCTP, and 1 μ l ddGTP was used. The fluorescent ddNTPs were

obtained from Amersham (Thermo Sequenase Dye Terminator Cycle Sequencing Core Kit, US 79565; Amersham did not publish the concentrations of the fluorescent nucleotides). The DNA fragment containing SNP HC21S00027 was digested with the restriction enzyme BsmF I, which generated a 5' overhang of four bases. As shown in FIG. 7E, if the first nucleotide incorporated is a labeled ddNTP, the 3' recessed end is filled in by one base, allowing detection of the SNP or locus of interest. However, if the first nucleotide incorporated is a dNTP, the polymerase continues to incorporate nucleotides until a ddNTP is filled in. For example, the first two nucleotides may be filled in with dNTPs, and the third nucleotide with a ddNTP, allowing detection of the third nucleotide in the overhang. Thus, the sequence of the entire 5' overhang may be determined, which increases the information obtained from each SNP or locus of interest.

[0289] After labeling, each Streptawell was rinsed with 1X PBS (100 μ l) three times. The "filled in" DNA fragments were then released from the Streptawells by digestion with the restriction enzyme EcoRI, according to the manufacturer's instructions that were supplied with the enzyme (FIGS. 8A-8D). Digestion was performed for 1 hour at 37 °C with shaking at 120 rpm.

Detection of the Locus of Interest

[0290] After release from the streptavidin matrix, 2-3 μ l of the 10 μ l sample was loaded in a 48 well membrane tray (The Gel Company, catalog number TAM48-01). The sample in the tray was absorbed with a 48 Flow Membrane Comb (The Gel Company, catalog number AM48), and inserted into a 36 cm 5% acrylamide (urea) gel (BioWhittaker Molecular Applications, Long Ranger Run Gel Packs, catalog number 50691).

[0291] The sample was electrophoresed into the gel at 3000 volts for 3 min. The membrane comb was removed, and the gel was run for 3 hours on an ABI 377 Automated Sequencing Machine. The incorporated labeled nucleotide was detected by fluorescence.

[0292] As shown in FIG. 9A, from a sample of thirty six (36) individuals, one of two nucleotides, either adenine or guanine, was detected at SNP

HC21S00027. These are the two nucleotides reported to exist at SNP HC21S00027 (www.snp.schl.org/snpsearch.shtml). One of two nucleotides, either guanine or cytosine, was detected at SNP TSC0095512 (FIG. 9B). The same results were obtained whether the locus of interest was amplified with a second primer that contained a recognition site for BceA I or the second primer contained a recognition site for BsmF I.

[0293] As shown in FIG. 9C, one of two nucleotides was detected at SNP TSC0264580, which was either adenosine or cytosine. These are the two nucleotides reported for this SNP site (www.snp.schl.org/snpsearch.shtml). In addition, a thymidine was detected one base upstream of the locus of interest. In a sequence dependent manner, BsmF I cuts some DNA molecules at the 10/14 position and other DNA molecules, which have the same sequence, at the 11/15 position. When the restriction enzyme BsmF I cuts 11 nucleotides away on the sense strand and 15 nucleotides away on the antisense strand, the 3' recessed end is one base upstream of the SNP site. The sequence of SNP TSC0264580 indicated that the base immediately preceding the SNP site was a thymidine. The incorporation of a labeled ddNTP into this position generated a fragment one base smaller than the fragment that was cut at the 10/14 position. Thus, the DNA molecules cut at the 11/15 position provided identity information about the base immediately preceding the SNP site, and the DNA molecules cut at the 10/14 position provided identity information about the SNP site.

[0294] SNP HC21S00027 was amplified using a second primer that contained the recognition site for BsmF I. A mixture of labeled ddNTPs and unlabeled dNTPs was used to fill in the 5' overhang generated by digestion with BsmF I. If a dNTP was incorporated, the polymerase continued to incorporate nucleotides until a ddNTP was incorporated. A population of DNA fragments, each differing by one base, was generated, which allowed the full sequence of the overhang to be determined.

[0295] As seen in FIG. 9D, an adenosine was detected, which was complementary to the nucleotide (a thymidine) immediately preceding the SNP or

locus of interest. This nucleotide was detected because of the 11/15 cutting property of BsmF I, which is described in detail above. A guanine and an adenine were detected at the SNP site, which are the two nucleotides reported for this SNP site (FIG. 9A). The two nucleotides were detected at the SNP site because the molecular weights of the dyes differ, which allowed separation of the two nucleotides. The next nucleotide detected was a thymidine, which is complementary to the nucleotide immediately downstream of the SNP site. The next nucleotide detected was a guanine, which was complementary to the nucleotide two bases downstream of the SNP site. Finally, an adenine was detected, which was complementary to the third nucleotide downstream of the SNP site. Sequence information was obtained not only for the SNP site but for the nucleotide immediately preceding the SNP site and the next three nucleotides.

[0296] None of the loci of interest contained a mutation. However, if one of the loci of interest harbored a mutation including but not limited to a point mutation, insertion, deletion, translocation or any combination of said mutations, it could be identified by comparison to the consensus or published sequence. Comparison of the sequences attributed to each of the loci of interest to the native, non-disease related sequence of the gene at each locus of interest determines the presence or absence of a mutation in that sequence. The finding of a mutation in the sequence is then interpreted as the presence of the indicated disease, or a predisposition to develop the same, as appropriate, in that individual. The relative amounts of the mutated vs. normal or non-mutated sequence can be assessed to determine if the subject has one or two alleles of the mutated sequence, and thus whether the subject is a carrier, or whether the indicated mutation results in a dominant or recessive condition.

EXAMPLE 3

[0297] Four loci of interest from chromosome 1 and two loci of interest from chromosome 21 were amplified in separate PCR reactions, pooled together, and analyzed. The primers were designed so that each amplified locus of interest was a different size, which allowed detection of the loci of interest.

Preparation of Template DNA

[0298] The template DNA was prepared from a 5 ml sample of blood obtained by venipuncture from a human volunteer with informed consent. Template DNA was isolated using the QIAamp DNA Blood Midi Kit supplied by QIAGEN (Catalog number 51183). The template DNA was isolated as per instructions included in the kit. Template DNA was isolated from thirty-six human volunteers, and then pooled into a single sample for further analysis.

Design of Primers

[0299] SNP TSC 0087315 was amplified using the following primers:

First primer:

5' TTACAATGCATGAATTCATCTTGGTCTCTCAAAGTGC 3' (SEQ ID NO:15)

Second primer:

5' TGGACCATAAACGGCCAAAAACTGTAAG 3' (SEQ ID NO:16)

[0300] SNP TSC0214366 was amplified using the following primers:

First primer:

5' ATGACTAGCTATGAATTGTTCAAGGTAGAAAATGGAA 3' (SEQ ID NO:13)

Second primer:

5' GAGAATTAGAACGGCCAAATCCCACTC 3' (SEQ ID NO:14)

[0301] SNP TSC 0413944 was amplified with the following primers:

First primer:

5' TACCTTTGATCGAATTCAAGGCCAAAATATTAAGTT 3' (SEQ ID NO:23)

Second primer:

5' TCGAACTTTAACGGCCTTAGAGTAGAGA 3' (SEQ ID NO:24)

[0302] SNP TSC0095512 was amplified using the following primers:

First primer:

5' AAGTTTAGATCAGAATTGAAAGCAGAAGTTGTCTG 3' (SEQ ID NO:11)

Second primer:

5' TCTCCAACTAACGGCTCATCGAGTAAAG 3' (SEQ ID NO:12)

[0303] SNP HC21S00131 was amplified with the following primers:

First primer:

5' CGATTCGATAAGAATTCAAAAGCAGTTCTAGTCAG 3' (SEQ ID NO:25)

Second primer:

5' TCGAATCTTACGGCTGCATCACATTCA 3' (SEQ ID NO:26)

[0304] SNP HC21S00027 was amplified with the following primers:

First primer:

5' ATAACCGTATGCGAATTCTATAATTCCTGATAAAGG 3' (SEQ ID NO:17)

Second primer:

5' CTTAAATCAGACGGCTAGGTAAACTTCA 3' (SEQ ID NO:19)

[0305] For each SNP, the first primer contained a recognition site for the restriction enzyme EcoRI and had a biotin tag at the extreme 5' end. The second primer used to amplify each SNP contained a recognition site for the restriction enzyme BceA I.

PCR Reaction

[0306] The PCR reactions were performed as described in Example 2 except that the following annealing temperatures were used: the annealing temperature for the first cycle of PCR was 37°C for 30 seconds, the annealing temperature for the second cycle of PCR was 57°C for 30 seconds, and the annealing temperature for the third cycle of PCR was 64°C for 30 seconds. All subsequent cycles had an annealing temperature of 64°C for 30 seconds. Thirty seven (37) cycles of PCR were performed. After PCR, 1/4 of the volume was removed from each reaction, and combined into a single tube.

Purification of Fragment Containing Locus of Interest

[0307] The PCR products (now combined into one sample, and referred to as "the sample") were separated from the genomic template DNA as described in Example 2 except that the sample was bound to a single well of a Streptawell microtiter plate.

Restriction Enzyme Digestion of Isolated Fragments Containing Loci of Interest

[0308] The sample was digested with the restriction enzyme BceA I, which bound the recognition site in the second primer. The restriction enzyme digestions were performed following the instructions supplied with the enzyme. After the restriction enzyme digest, the wells were washed three times with 1X PBS.

Incorporation of Nucleotides

[0309] The restriction enzyme digest described above yielded DNA molecules with a 5' overhang, which contained the SNP site or locus of interest and a 3' recessed end. The 5' overhang functioned as a template allowing incorporation of a nucleotide in the presence of a DNA polymerase.

[0310] The following components were used for the fill in reaction: 1 μ l of fluorescently labeled ddATP; 1 μ l of fluorescently labeled ddTTP; 1 μ l of fluorescently labeled ddGTP; 1 μ l of fluorescently labeled ddCTP; 2 μ l of 10X sequenase buffer, 0.25 μ l of Sequenase, and water as needed for a 20 μ l reaction. The fill in reaction was performed at 40°C for 10 min. All labeling reagents were obtained from Amersham (Thermo Sequenase Dye Terminator Cycle Sequencing Core Kit (US 79565); the concentration of the ddNTPS provided in the kit is proprietary and not published by Amersham). In the presence of fluorescently labeled ddNTPs, the 3' recessed end was filled in by one base, which corresponds to the SNP or locus of interest.

[0311] After the incorporation of nucleotide, the Streptawell was rinsed with 1X PBS (100 μ l) three times. The "filled in" DNA fragments were then released from the Streptawell by digestion with the restriction enzyme EcoRI following the

manufacturer's instructions. Digestion was performed for 1 hour at 37 °C with shaking at 120 rpm.

Detection of the Locus of Interest

[0312] After release from the streptavidin matrix, 2-3 µl of the 10 µl sample was loaded in a 48 well membrane tray (The Gel Company, catalog number TAM48-01). The sample in the tray was absorbed with a 48 Flow Membrane Comb (The Gel Company, catalog number AM48), and inserted into a 36 cm 5% acrylamide (urea) gel (BioWhittaker Molecular Applications, Long Ranger Run Gel Packs, catalog number 50691).

[0313] The sample was electrophoresed into the gel at 3000 volts for 3 min. The membrane comb was removed, and the gel was run for 3 hours on an ABI 377 Automated Sequencing Machine. The incorporated nucleotide was detected by fluorescence.

[0314] The primers were designed so that each amplified locus of interest differed in size. As shown in FIG. 10, each amplified loci of interest differed by about 5-10 nucleotides, which allowed the loci of interest to be separated from one another by gel electrophoresis. Two nucleotides were detected for SNP TSC0087315, which were guanine and cytosine. These are the two nucleotides reported to exist at SNP TSC0087315 (www.snp.schl.org/snpsearch.shtml). The sample comprised template DNA from 36 individuals and because the DNA molecules that incorporated a guanine differed in molecular weight from those that incorporated a cytosine, distinct bands were seen for each nucleotide.

[0315] Two nucleotides were detected at SNP HC21S00027, which were guanine and adenine (FIG. 10). The two nucleotides reported for this SNP site are guanine and adenine (www.snp.schl.org/snpsearch.shtml). As discussed above, the sample contained template DNA from thirty-six individuals, and one would expect both nucleotides to be represented in the sample. The molecular weight of the DNA fragments that incorporated a guanine was distinct from the DNA fragments that incorporated an adenine, which allowed both nucleotides to be detected.

[0316] The nucleotide cytosine was detected at SNP TSC0214366 (FIG. 10).

The two nucleotides reported to exist at this SNP position are thymidine and cytosine.

[0317] The nucleotide guanine was detected at SNP TSC0413944 (FIG. 10).

The two nucleotides reported for this SNP are guanine and cytosine (<http://snp.cshl.org/snpsearch.shtml>).

[0318] The nucleotide cytosine was detected at SNP TSC0095512 (FIG. 10).

The two nucleotides reported for this SNP site are guanine and cytosine ([www.snp.cshl.org/snpsearch.shtml](http://snp.cshl.org/snpsearch.shtml)).

[0319] The nucleotide detected at SNP HC21S00131 was guanine. The two nucleotides reported for this SNP site are guanine and adenosine ([www.snp.cshl.org/snpsearch.shtml](http://snp.cshl.org/snpsearch.shtml)).

[0320] As discussed above, the sample was comprised of DNA templates from thirty-six individuals and one would expect both nucleotides at the SNP sites to be represented. For SNP TSC0413944, TSC0095512, TSC0214366 and HC21S00131, one of the two nucleotides was detected. It is likely that both nucleotides reported for these SNP sites are present in the sample but that one fluorescent dye overwhelms the other. The molecular weight of the DNA molecules that incorporated one nucleotide did not allow efficient separation of the DNA molecules that incorporated the other nucleotide. However, the SNPs were readily separated from one another, and for each SNP, a proper nucleotide was incorporated. The sequences of multiple loci of interest from multiple chromosomes, which were treated as a single sample after PCR, were determined.

[0321] A single reaction containing fluorescently labeled ddNTPs was performed with the sample that contained multiple loci of interest. Alternatively, four separate fill in reactions can be performed where each reaction contains one fluorescently labeled nucleotide (ddATP, ddTTP, ddGTP, or ddCTP) and unlabeled ddNTPs (see Example 2, FIGS. 7A-7D and FIGS. 9A-C). Four separate "fill in" reactions will allow detection of any nucleotide that is present at the loci of interest. For example, if analyzing a sample that contains multiple loci

of interest from a single individual, and said individual is heterozygous at one or more than one loci of interest, four separate "fill in" reactions can be used to determine the nucleotides at the heterozygous loci of interest.

[0322] Also, when analyzing a sample that contains templates from multiple individuals, four separate "fill in" reactions will allow detection of nucleotides present in the sample, independent of how frequent the nucleotide is found at the locus of interest. For example, if a sample contains DNA templates from 50 individuals, and 49 of the individuals have a thymidine at the locus of interest, and one individual has a guanine, the performance of four separate "fill in" reactions, wherein each "fill in" reaction is run in a separate lane of a gel, such as in FIGS. 9A-9C, will allow detection of the guanine. When analyzing a sample comprised of multiple DNA templates, multiple "fill in" reactions will alleviate the need to distinguish multiple nucleotides at a single site of interest by differences in mass.

[0323] In this example, multiple single nucleotide polymorphisms were analyzed. It is also possible to determine the presence or absence of mutations, including point mutations, transitions, transversions, translocations, insertions, and deletions from multiple loci of interest. The multiple loci of interest can be from a single chromosome or from multiple chromosomes. The multiple loci of interest can be from a single gene or from multiple genes.

[0324] The sequence of multiple loci of interest that cause or predispose to a disease phenotype can be determined. For example, one could amplify one to tens to hundreds to thousands of genes implicated in cancer or any other disease. The primers can be designed so that each amplified loci of interest differs in size. After PCR, the amplified loci of interest can be combined and treated as a single sample. Alternatively, the multiple loci of interest can be amplified in one PCR reaction or the total number of loci of interest, for example 100, can be divided into samples, for example 10 loci of interest per PCR reaction, and then later pooled. As demonstrated herein, the sequence of multiple loci of interest can be determined. Thus, in one reaction, the sequence of one to ten to hundreds to

thousands of genes that predispose or cause a disease phenotype can be determined.

EXAMPLE 4

[0325] Genomic DNA was obtained from four individuals after informed consent was obtained. Six SNPs on chromosome 13 (TSC0837969, TSC0034767, TSC1130902, TSC0597888, TSC0195492, TSC0607185) were analyzed using the template DNA. Information regarding these SNPs can be found at the following website (www.snp.schl.org/snpsearch.shtml)website active as of February 11, 2003).

[0326] A single nucleotide labeled with one fluorescent dye was used to genotype the individuals at the six selected SNP sites. The primers were designed to allow the six SNPs to be analyzed in a single reaction.

Preparation of Template DNA

[0327] The template DNA was prepared from a 9 ml sample of blood obtained by venipuncture from a human volunteer with informed consent. Template DNA was isolated using the QIAamp DNA Blood Midi Kit supplied by QIAGEN (Catalog number 51183). The template DNA was isolated as per instructions included in the kit.

Design of Primers

[0328] SNP TSC0837969 was amplified using the following primer set:

First primer:

5' GGGCTAGTCTCCGAATTCCACCTATCCTACCAAATGTC 3'

Second primer:

5' TAGCTGTAGTTAGGGACTGTTCTGAGCAC 3'

[0329] The first primer had a biotin tag at the 5' end and contained a restriction enzyme recognition site for EcoRI. The first primer was designed to anneal 44 bases from of the locus of interest. The second primer contained a restriction enzyme recognition site for BsmF I.

[0330] SNP TSC0034767 was amplified using the following primer set:

First primer:

5' CGAATGCAAGGCGAATTCTGTTAGTAATAACACAGTGCA 3'

Second primer:

5' AAGACTGGATCCGGGACCATGTAGAATAC 3'

[0331] The first primer had a biotin tag at the 5' end and contained a restriction enzyme recognition site for EcoRI. The first primer was designed to anneal 50 bases from the locus of interest. The second primer contained a restriction enzyme recognition site for BsmF I.

[0332] SNP TSC1130902 was amplified using the following primer set:

First primer:

5' TCTAACCAATTGCGAATTCAAGGGCAAGGGGGGTGAGATC 3'

Second primer:

5' TGACTTGGATCCGGGACAACGACTCATCC 3'

[0333] The first primer had a biotin tag at the 5' end and contained a restriction enzyme recognition site for EcoRI. The first primer was designed to anneal 60 bases from the locus of interest. The second primer contained a restriction enzyme recognition site for BsmF I.

[0334] SNP TSC0597888 was amplified using the following primer set:

First primer:

5' ACCCAGGCGCCAGAATTCTTAGATAAAGCTGAAGGGA 3'

Second primer:

5' GTTACGGGATCCGGGACTCCATATTGATC 3'

[0335] The first primer had a biotin tag at the 5' end and contained a restriction enzyme recognition site for EcoRI. The first primer was designed to anneal 70 bases from the locus of interest. The second primer contained a restriction enzyme recognition site for BsmF I.

[0336] SNP TSC0195492 was amplified using the following primer set:

First primer:

5' CGTTGGCTTGAGGAATTGACCAAAAGAGCCAAGAGAA

Second primer:

5' AAAAAGGGATCCGGGACCTTGACTAGGAC 3'

[0337] The first primer had a biotin tag at the 5' end and contained a restriction enzyme recognition site for EcoRI. The first primer was designed to anneal 80 bases from the locus of interest. The second primer contained a restriction enzyme recognition site for BsmF I.

[0338] SNP TSC0607185 was amplified using the following primer set:

First primer:

5' ACTTGATTCCGTGAATTCTTATCAATAATCTTACAT 3'

Second primer:

5' CAAGTTGGATCCGGGACCCAGGGCTAACCC 3'

[0339] The first primer had a biotin tag at the 5' end and contained a restriction enzyme recognition site for EcoRI. The first primer was designed to anneal 90 bases from the locus of interest. The second primer contained a restriction enzyme recognition site for BsmF I.

[0340] All loci of interest were amplified from the template genomic DNA using the polymerase chain reaction (PCR, U.S. Patent Nos. 4,683,195 and 4,683,202, incorporated herein by reference). In this example, the loci of interest were amplified in separate reaction tubes but they could also be amplified together in a single PCR reaction. For increased specificity, a "hot-start" PCR was used. PCR reactions were performed using the HotStarTaq Master Mix Kit supplied by QIAGEN (catalog number 203443). The amount of template DNA and primer per reaction can be optimized for each locus of interest but in this example, 40 ng of template human genomic DNA and 5 μ M of each primer were used. Forty cycles of PCR were performed. The following PCR conditions were used:

- (1) 95°C for 15 minutes and 15 seconds;
- (2) 37°C for 30 seconds;
- (3) 95°C for 30 seconds;
- (4) 57°C for 30 seconds;
- (5) 95°C for 30 seconds;
- (6) 64°C for 30 seconds;

- (7) 95°C for 30 seconds;
- (8) Repeat steps 6 and 7 thirty nine (39) times;
- (9) 72°C for 5 minutes.

[0341] In the first cycle of PCR, the annealing temperature was about the melting temperature of the 3' annealing region of the second primers, which was 37°C. The annealing temperature in the second cycle of PCR was about the melting temperature of the 3' region, which anneals to the template DNA, of the first primer, which was 57°C. The annealing temperature in the third cycle of PCR was about the melting temperature of the entire sequence of the second primer, which was 64°C. The annealing temperature for the remaining cycles was 64°C. Escalating the annealing temperature from TM1 to TM2 to TM3 in the first three cycles of PCR greatly improves specificity. These annealing temperatures are representative, and the skilled artisan will understand the annealing temperatures for each cycle are dependent on the specific primers used.

[0342] The temperatures and times for denaturing, annealing, and extension, can be optimized by trying various settings and using the parameters that yield the best results. In this example, the first primer was designed to anneal at various distances from the locus of interest. The skilled artisan understands that the annealing location of the first primer can be 5-10, 11-15, 16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-65, 66-70, 71-75, 76-80, 81-85, 86-90, 91-95, 96-100, 101-105, 106-110, 111-115, 116-120, 121-125, 126-130, 131-140, 141-160, 161-180, 181-200, 201-220, 221-240, 241-260, 261-280, 281-300, 301-350, 351-400, 401-450, 451-500, or greater than 500 bases from the locus of interest.

Purification of Fragment Containing Locus of Interest

[0343] The PCR products were separated from the genomic template DNA. After the PCR reaction, 1/4 of the volume of each PCR reaction from one individual was mixed together in a well of a Streptawell, transparent, High-Bind plate from Roche Diagnostics GmbH (catalog number 1 645 692, as listed in Roche Molecular Biochemicals, 2001 Biochemicals Catalog). The first primers

contained a 5' biotin tag so the PCR products bound to the Streptavidin coated wells while the genomic template DNA did not. The streptavidin binding reaction was performed using a Thermomixer (Eppendorf) at 1000 rpm for 20 min. at 37°C. Each well was aspirated to remove unbound material, and washed three times with 1X PBS, with gentle mixing (Kandpal et al., Nucl. Acids Res. 18:1789-1795 (1990); Kaneoka et al., Biotechniques 10:30-34 (1991); Green et al., Nucl. Acids Res. 18:6163-6164 (1990)).

Restriction Enzyme Digestion of Isolated Fragments Containing Loci of Interest [0344] The purified PCR products were digested with the restriction enzyme BsmF I, which binds to the recognition site incorporated into the PCR products from the second primer. The digests were performed in the Streptawells following the instructions supplied with the restriction enzyme. After digestion, the wells were washed three times with PBS to remove the cleaved fragments.

Incorporation of Labeled Nucleotide

[0345] The restriction enzyme digest with BsmF I yielded a DNA fragment with a 5' overhang, which contained the SNP site or locus of interest and a 3' recessed end. The 5' overhang functioned as a template allowing incorporation of a nucleotide or nucleotides in the presence of a DNA polymerase.

[0346] Below, a schematic of the 5' overhang for SNP TSC0837969 is shown. The entire DNA sequence is not reproduced, only the portion to demonstrate the overhang (where R indicates the variable site).

| | | | | |
|-------------------|---|---|---|---|
| 5' TTAA | | | | |
| 3' AATT | R | A | C | A |
| Overhang position | 1 | 2 | 3 | 4 |

[0347] The observed nucleotides for TSC0837969 on the 5' sense strand (here depicted as the top strand) are adenine and guanine. The third position in the overhang on the antisense strand corresponds to cytosine, which is complementary to guanine. As this variable site can be adenine or guanine, fluorescently labeled ddGTP in the presence of unlabeled dCTP, dTTP, and dATP

was used to determine the sequence of both alleles. The fill-in reactions for an individual homozygous for guanine, homozygous for adenine or heterozygous are diagrammed below.

[0348] **Homozygous for guanine at TSC 0837969:**

| | | | | | |
|-------------------|---------|----|---|---|---|
| Allele 1 | 5' TTAA | G* | | | |
| | 3' AATT | C | A | C | A |
| Overhang position | | 1 | 2 | 3 | 4 |
| Allele 2 | 5' TTAA | G* | | | |
| | 3' AATT | C | A | C | A |
| Overhang position | | 1 | 2 | 3 | 4 |

[00100] Labeled ddGTP is incorporated into the first position of the overhang. Only one signal is seen, which corresponds to the molecules filled in with labeled ddGTP at the first position of the overhang.

[0349] **Homozygous for adenine at TSC 0837969:**

| | | | | | |
|-------------------|---------|---|---|----|---|
| Allele 1 | 5' TTAA | A | T | G* | |
| | 3' AATT | T | A | C | A |
| Overhang position | | 1 | 2 | 3 | 4 |
| Allele 2 | 5' TTAA | A | T | G* | |
| | 3' AATT | T | A | C | A |
| Overhang position | | 1 | 2 | 3 | 4 |

[0350] Unlabeled dATP is incorporated at position one of the overhang, and unlabeled dTTP is incorporated at position two of the overhang. Labeled ddGTP is incorporated at position three of the overhang. Only one signal will be seen; the molecules filled in with ddGTP at position 3 will have a different molecular weight from molecules filled in at position one, which allows easy identification of individuals homozygous for adenine or guanine.

[0351] **Heterozygous at TSC0837969:**

| | | | | | |
|-------------------|---------|----|---|---|---|
| Allele 1 | 5' TTAA | G* | | | |
| | 3' AATT | C | A | C | A |
| Overhang position | | 1 | 2 | 3 | 4 |

| | | | | |
|-------------------|---------|---|---|-----|
| Allele 2 | 5' TTAA | A | T | G* |
| | 3' AATT | T | A | C A |
| Overhang position | | 1 | 2 | 3 4 |

[0352] Two signals will be seen; one signal corresponds to the DNA molecules filled in with ddGTP at position 1, and a second signal corresponding to molecules filled in at position 3 of the overhang. The two signals can be separated using any technique that separates based on molecular weight including but not limited to gel electrophoresis.

[0353] Below, a schematic of the 5' overhang for SNP TSC0034767 is shown. The entire DNA sequence is not reproduced, only the portion to demonstrate the overhang (where R indicates the variable site).

| | | | | |
|---|---|---|---|-------------------|
| A | C | A | R | GTGT 3' |
| | | | | CACA 5' |
| 4 | 3 | 2 | 1 | Overhang Position |

[0354] The observed nucleotides for TSC0034767 on the 5' sense strand (here depicted as the top strand) are cytosine and guanine. The second position in the overhang corresponds to adenine, which is complementary to thymidine. The third position in the overhang corresponds to cytosine, which is complementary to guanine. Fluorescently labeled ddGTP in the presence of unlabeled dCTP, dTTP, and dATP is used to determine the sequence of both alleles.

[0355] In this case, the second primer anneals upstream of the locus of interest, and thus the fill-in reaction occurs on the anti-sense strand (here depicted as the bottom strand). Either the sense strand or the antisense strand can be filled in depending on whether the second primer, which contains the type IIS restriction enzyme recognition site, anneals upstream or downstream of the locus of interest.

Below, a schematic of the 5' overhang for SNP TSC1130902 is shown. The entire DNA sequence is not reproduced, only a portion to demonstrate the overhang (where R indicates the variable site).

5' TTCAT

| | | | | |
|-------------------|---|---|---|---|
| 3' AAGTA | R | T | C | C |
| Overhang position | 1 | 2 | 3 | 4 |

[00101] The observed nucleotides for TSC1130902 on the 5' sense strand are adenine and guanine. The second position in the overhang corresponds to a thymidine, and the third position in the overhang corresponds to cytosine, which is complementary to guanine.

[0356] Fluorescently labeled ddGTP in the presence of unlabeled dCTP, dTTP, and dATP is used to determine the sequence of both alleles.

[00102] Below, a schematic of the 5' overhang for SNP TSC0597888 is shown. The entire DNA sequence is not reproduced, only the portion to demonstrate the overhang (where R indicates the variable site).

| | | | | |
|---|---|---|---|-------------------|
| T | C | T | R | ATTC 3' |
| | | | | TAAG 5' |
| 4 | 3 | 2 | 1 | Overhang position |

[0357] The observed nucleotides for TSC0597888 on the 5' sense strand (here depicted as the top strand) are cytosine and guanine. The third position in the overhang corresponds to cytosine, which is complementary to guanine.

Fluorescently labeled ddGTP in the presence of unlabeled dCTP, dTTP, and dATP is used to determine the sequence of both alleles.

[0358] Below, a schematic of the 5' overhang for SNP TSC0607185 is shown. The entire DNA sequence is not reproduced, only the portion to demonstrate the overhang (where R indicates the variable site).

| | | | | |
|---|---|---|---|-------------------|
| C | C | T | R | TGTC 3' |
| | | | | ACAG 5' |
| 4 | 3 | 2 | 1 | Overhang position |

[0359] The observed nucleotides for TSC0607185 on the 5' sense strand (here depicted as the top strand) are cytosine and thymidine. In this case, the second primer anneals upstream of the locus of interest, which allows the anti-sense strand to be filled in. The anti-sense strand (here depicted as the bottom strand) will be filled in with guanine or adenine.

[0360] The second position in the 5' overhang is thymidine, which is complementary to adenine, and the third position in the overhang corresponds to cytosine, which is complementary to guanine. Fluorescently labeled ddGTP in the presence of unlabeled dCTP, dTTP, and dATP is used to determine the sequence of both alleles.

[0361] Below, a schematic of the 5' overhang for SNP TSC0195492 is shown. The entire DNA sequence is not reproduced, only the portion to demonstrate the overhang.

| | | | | |
|-------------------|---|---|---|---|
| 5' ATCT | | | | |
| 3' TAGA | R | A | C | A |
| Overhang position | 1 | 2 | 3 | 4 |

[0362] The observed nucleotides at this site are cytosine and guanine on the sense strand (here depicted as the top strand). The second position in the 5' overhang is adenine, which is complementary to thymidine, and the third position in the overhang corresponds to cytosine, which is complementary to guanine. Fluorescently labeled ddGTP in the presence of unlabeled dCTP, dTTP, and dATP was used to determine the sequence of both alleles.

[0363] As demonstrated above, the sequence of both alleles of the six SNPs can be determined by labeling with ddGTP in the presence of unlabeled dATP, dTTP, and dCTP. The following components were added to each fill in reaction: 1 μ l of fluorescently labeled ddGTP, 0.5 μ l of unlabeled ddNTPs (40 μ M), which contained all nucleotides except guanine, 2 μ l of 10X sequenase buffer, 0.25 μ l of Sequenase, and water as needed for a 20 μ l reaction. The fill in reaction was performed at 40°C for 10 min. Non-fluorescently labeled ddNTP was purchased from Fermentas Inc. (Hanover, MD). All other labeling reagents were obtained from Amersham (Thermo Sequenase Dye Terminator Cycle Sequencing Core Kit, US 79565).

[0364] After labeling, each Streptawell was rinsed with 1X PBS (100 μ l) three times. The "filled in" DNA fragments were then released from the Streptawells by digestion with the restriction enzyme EcoRI, according to the manufacturer's

instructions that were supplied with the enzyme. Digestion was performed for 1 hour at 37 °C with shaking at 120 rpm.

Detection of the Locus of Interest

[0365] After release from the streptavidin matrix, the sample was loaded into a lane of a 36 cm 5% acrylamide (urea) gel (BioWhittaker Molecular Applications, Long Ranger Run Gel Packs, catalog number 50691). The sample was electrophoresed into the gel at 3000 volts for 3 min. The gel was run for 3 hours on a sequencing apparatus (Hoefer SQ3 Sequencer). The gel was removed from the apparatus and scanned on the Typhoon 9400 Variable Mode Imager. The incorporated labeled nucleotide was detected by fluorescence.

[0366] As shown in FIG. 11, the template DNA in lanes 1 and 2 for SNP TSC0837969 is homozygous for adenine. The following fill-in reaction was expected to occur if the individual was homozygous for adenine:

[0367] Homozygous for adenine at TSC 0837969:

| | | | |
|-------------------|---|---|-----|
| 5' TTAA | A | T | G* |
| 3' AATT | T | A | C A |
| Overhang position | 1 | 2 | 3 4 |

[0368] Unlabeled dATP was incorporated in the first position complementary to the overhang. Unlabeled dTTP was incorporated in the second position complementary to the overhang. Labeled ddGTP was incorporated in the third position complementary to the overhang. Only one band was seen, which migrated at about position 46 of the acrylamide gel. This indicated that adenine was the nucleotide filled in at position one. If the nucleotide guanine had been filled in, a band would be expected at position 44.

[0369] However, the template DNA in lanes 3 and 4 for SNP TSC0837969 was heterozygous. The following fill-in reactions were expected if the individual was heterozygous:

[0370] Heterozygous at TSC0837969:

| | | |
|----------|---------|---------|
| Allele 1 | 5' TTAA | G* |
| | 3' AATT | C A C A |

| | | | | |
|-----------------------|---|---|----|---|
| Overhang position | 1 | 2 | 3 | 4 |
| Allele 2 5' TTAA | A | T | G* | |
| 3' AATT | T | A | C | A |
| Overhang position | 1 | 2 | 3 | 4 |

[0371] Two distinct bands were seen; one band corresponds to the molecules filled in with ddGTP at position 1 complementary to the overhang (the G allele), and the second band corresponds to molecules filled in with ddGTP at position 3 complementary to the overhang (the A allele). The two bands were separated based on the differences in molecular weight using gel electrophoresis. One fluorescently labeled nucleotide ddGTP was used to determine that an individual was heterozygous at a SNP site. This is the first use of a single nucleotide to effectively detect the presence of two different alleles.

[0372] For SNP TSC0034767, the template DNA in lanes 1 and 3 is heterozygous for cytosine and guanine, as evidenced by the two distinct bands. The lower band corresponds to ddGTP filled in at position 1 complementary to the overhang. The second band of slightly higher molecular weight corresponds to ddGTP filled in at position 3, indicating that the first position in the overhang was filled in with unlabeled dCTP, which allowed the polymerase to continue to incorporate nucleotides until it incorporated ddGTP at position 3 complementary to the overhang. The template DNA in lanes 2 and 4 was homozygous for guanine, as evidenced by a single band of higher molecular weight than if ddGTP had been filled in at the first position complementary to the overhang.

[0373] For SNP TSC1130902, the template DNA in lanes 1, 2, and 4 is homozygous for adenine at the variable site, as evidenced by a single higher molecular weight band migrating at about position 62 on the gel. The template DNA in lane 3 is heterozygous at the variable site, as indicated by the presence of two distinct bands. The lower band corresponded to molecules filled in with ddGTP at position 1 complementary to the overhang (the guanine allele). The higher molecular weight band corresponded to molecules filled in with ddGTP at position 3 complementary to the overhang (the adenine allele).

[0374] For SNP TSC0597888, the template DNA in lanes 1 and 4 was homozygous for cytosine at the variable site; the template DNA in lane 2 was heterozygous at the variable site, and the template DNA in lane 3 was homozygous for guanine. The expected fill-in reactions are diagrammed below:

[0375] Homozygous for cytosine:

| | | | | | |
|----------|---|----|---|---|-------------------|
| Allele 1 | T | C | T | G | ATTC 3' |
| | | G* | A | C | TAAG 5' |
| 4 | 3 | 2 | 1 | | Overhang position |

| | | | | | |
|----------|---|----|---|---|-------------------|
| Allele 2 | T | C | T | G | ATTC 3' |
| | | G* | A | C | TAAG 5' |
| 4 | 3 | 2 | 1 | | Overhang position |

[0376] Homozygous for guanine:

| | | | | | |
|----------|---|---|---|----|-------------------|
| Allele 1 | T | C | T | C | ATTC 3' |
| | | | | G* | TAAG 5' |
| 4 | | 3 | 2 | 1 | Overhang position |

| | | | | | |
|----------|---|---|---|----|-------------------|
| Allele 2 | T | C | T | C | ATTC 3' |
| | | | | G* | TAAG 5' |
| 4 | | 3 | 2 | 1 | Overhang position |

[0377] Heterozygous for guanine/cytosine:

| | | | | | |
|----------|---|----|---|---|-------------------|
| Allele 1 | T | C | T | G | ATTC 3' |
| | | G* | A | C | TAAG 5' |
| 4 | | 3 | 2 | 1 | Overhang position |

| | | | | | |
|----------|---|---|---|---|---------|
| Allele 2 | T | C | T | C | ATTC 3' |
|----------|---|---|---|---|---------|

| | | | | |
|---|--|---|----|---------------------|
| | | | G* | TAAG 5' |
| 4 | | 3 | 2 | 1 Overhang position |

[0378] Template DNA homozygous for guanine at the variable site displayed a single band, which corresponded to the DNA molecules filled in with ddGTP at position 1 complementary to the overhang. These DNA molecules were of lower molecular weight compared to the DNA molecules filled in with ddGTP at position 3 of the overhang (see lane 3 for SNP TSC0597888). The DNA molecules differed by two bases in molecular weight.

[0379] Template DNA homozygous for cytosine at the variable site displayed a single band, which corresponds to the DNA molecules filled in with ddGTP at position 3 complementary to the overhang. These DNA molecules migrated at a higher molecular weight than DNA molecules filled in with ddGTP at position 1 (see lanes 1 and 4 for SNP TSC0597888).

[0380] Template DNA heterozygous at the variable site displayed two bands; one band corresponded to the DNA molecules filled in with ddGTP at position 1 complementary to the overhang and was of lower molecular weight, and the second band corresponded to DNA molecules filled in with ddGTP at position 3 complementary to the overhang, and was of higher molecular weight (see lane 3 for SNP TSC0597888).

[0381] For SNP TSC0195492, the template DNA in lanes 1 and 3 was heterozygous at the variable site, which was demonstrated by the presence of two distinct bands. The template DNA in lane 2 was homozygous for guanine at the variable site. The template DNA in lane 4 was homozygous for cytosine. Only one band was seen in lane 4 for this SNP, and it had a higher molecular weight than the DNA molecules filled in with ddGTP at position 1 complementary to the overhang (compare lanes 2, 3 and 4).

[0382] The observed alleles for SNP TSC0607185 are reported as cytosine or thymidine. For consistency, the SNP consortium denotes the observed alleles as they appear in the sense strand (www.snp.schl.org/snpsearch.shtml); website

active as of February 11, 2003). For this SNP, the second primer annealed upstream of the locus of interest, which allowed the fill-in reaction to occur on the antisense strand after digestion with BsmF I.

[0383] The template DNA in lanes 1 and 3 was heterozygous; the template DNA in lane 2 was homozygous for thymidine, and the template DNA in lane 4 was homozygous for cytosine. The antisense strand was filled in with ddGTP, so the nucleotide on the sense strand corresponded to cytosine.

[0384] Molecular weight markers can be used to identify the positions of the expected bands. Alternatively, for each SNP analyzed, a known heterozygous sample can be used, which will identify precisely the position of the two expected bands.

[0385] As demonstrated in FIG. 11, one nucleotide labeled with one fluorescent dye can be used to determine the identity of a variable site including but not limited to SNPs and single nucleotide mutations. Typically, to determine if an individual is homozygous or heterozygous at a SNP site, multiple reactions are performed using one nucleotide labeled with one dye and a second nucleotide labeled with a second dye. However, this introduces problems in comparing results because the two dyes have different quantum coefficients. Even if different nucleotides are labeled with the same dye, the quantum coefficients are different. The use of a single nucleotide labeled with one dye eliminates any errors from the quantum coefficients of different dyes.

[0386] In this example, fluorescently labeled ddGTP was used. However, the method is applicable for a nucleotide tagged with any signal generating moiety including but not limited to radioactive molecule, fluorescent molecule, antibody, antibody fragment, hapten, carbohydrate, biotin, derivative of biotin, phosphorescent moiety, luminescent moiety, electrochemiluminescent moiety, chromatic moiety, and moiety having a detectable electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity. In addition, labeled ddATP, ddTTP, or ddCTP can be used.

[0387] The above example used the third position complementary to the overhang as an indicator of the second allele. However, the second or fourth position of the overhang can be used as well (see Section on Incorporation of Nucleotides). Furthermore, the overhang was generated with the type IIS enzyme BsmF I; however any enzyme that cuts DNA at a distance from its binding site can be used including but not limited to the enzymes listed in Table I.

[0388] Also, in the above example, the nucleotide immediately preceding the SNP site was not a guanine on the strand that was filled in. This eliminated any effects of the alternative cutting properties of the type IIS restriction enzyme to be removed. For example, at SNP TSC0837969, the nucleotide upstream of the SNP site on the sense strand was an adenine. If BsmF I displayed alternate cutting properties, the following overhangs would be generated for the adenine allele and the guanine allele:

| | | | | | |
|------------------------|--------|---|-----------|---|-----------|
| G allele – 11/15 Cut | 5' TTA | | | | |
| | 3' AAT | T | C | A | C |
| Overhang position | | 0 | 1 | 2 | 3 |
| | | | | | |
| G allele after fill-in | 5' TTA | A | G* | | |
| | 3' AAT | T | C | A | C |
| Overhang position | | 0 | 1 | 2 | 3 |
| | | | | | |
| A allele 11/15 Cut | 5' TTA | | | | |
| | 3' AAT | T | T | A | C |
| Overhang position | | 0 | 1 | 2 | 3 |
| | | | | | |
| A allele after fill-in | 5' TTA | A | A | T | G* |
| | 3' AAT | T | T | A | C |
| Overhang position | | 0 | 1 | 2 | 3 |

[0389] For the guanine allele, the first position in the overhang would be filled in with dATP, which would allow the polymerase to incorporate ddGTP at position 2 complementary to the overhang. There would be no detectable difference between molecules cut at the 10/14 position or molecules cut at the 11/15 position.

[0390] For the adenine allele, the first position complementary to the overhang would be filled in with dATP, the second position would be filled in with dATP, the third position would be filled in with dTTP, and the fourth position would be filled in with ddGTP. There would be no difference in the molecular weights between molecules cut at 10/14 or molecules cut at 11/15. The only differences would correspond to whether the DNA molecules contained an adenine at the variable site or a guanine at the variable site.

[0391] As seen in FIG. 11, positioning the annealing region of the first primer allows multiple SNPs to be analyzed in a single lane of a gel. Also, when using the same nucleotide with the same dye, a single fill-in reaction can be performed. In this example, 6 SNPs were analyzed in one lane. However, any number of SNPs including but not limited to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-40, 41-50, 51-60, 61-70, 71-80, 81-100, 101-120, 121-140, 141-160, 161-180, 181-200, and greater than 200 can be analyzed in a single reaction.

[0392] Furthermore, one labeled nucleotide used to detect both alleles can be mixed with a second labeled nucleotide used to detect a different set of SNPs provided that neither of the nucleotides that are labeled occur immediately before the variable site (complementary to nucleotide at position 0 of the 11/15 cut). For example, suppose SNP X can be guanine or thymidine at the variable site and has the following 5' overhang generated after digestion with BsmF I:

| | | | | | | |
|-------------------|----------|---|---|---|---|--|
| SNP X 10/14 | 5' TTGAC | | | | | |
| G allele | 3' AACTG | C | A | C | T | |
| Overhang position | | 1 | 2 | 3 | 4 | |

| | | | | | |
|-------------------|---------|---|---|---|---|
| SNP X 11/15 | 5' TTGA | | | | |
| G allele | 3'AACT | G | C | A | C |
| Overhang position | | 0 | 1 | 2 | 3 |

| | | | | | |
|-------------------|----------|---|---|---|---|
| SNP X 10/14 | 5' TTGAC | | | | |
| T allele | 3'AACTG | A | A | C | T |
| Overhang position | | 1 | 2 | 3 | 4 |

| | | | | | |
|-------------------|---------|---|---|---|---|
| SNP X 11/15 | 5' TTGA | | | | |
| T allele | 3'AACT | G | A | A | C |
| Overhang position | | 0 | 1 | 2 | 3 |

[0393] After the fill-in reaction with labeled ddGTP, unlabeled dATP, dCTP, and dTTP, the following molecules would be generated:

| | | | | | |
|-------------------|----------|----|---|---|---|
| SNP X 10/14 | 5' TTGAC | G* | | | |
| G allele | 3'AACTG | C | A | C | T |
| Overhang position | | 1 | 2 | 3 | 4 |

| | | | | | |
|-------------------|---------|---|----|---|---|
| SNP X 11/15 | 5' TTGA | C | G* | | |
| G allele | 3'AACT | G | C | A | C |
| Overhang position | | 0 | 1 | 2 | 3 |

| | | | | | |
|-------------------|----------|---|---|----|---|
| SNP X 10/14 | 5' TTGAC | T | T | G* | |
| T allele | 3'AACTG | A | A | C | T |
| Overhang position | | 1 | 2 | 3 | 4 |

| | | | | | |
|-------------|---------|---|---|---|----|
| SNP X 11/15 | 5' TTGA | C | T | T | G* |
| T allele | 3'AACT | G | A | A | C |

| Overhang position | 0 | 1 | 2 | 3 |
|-------------------|---|---|---|---|
|-------------------|---|---|---|---|

[0394] Now suppose SNP Y can be adenine or thymidine and has the following 5' overhangs generated after digestion with BsmF I.

| | | | | | |
|-------------------|---------|---|---|---|---|
| SNP Y 10/14 | 5' GTTT | | | | |
| A allele | 3' CAAA | T | G | T | A |
| Overhang position | | 1 | 2 | 3 | 4 |

| | | | | | |
|-------------------|--------|---|---|---|---|
| SNP Y 11/15 | 5' GTT | | | | |
| A allele | 3' CAA | A | T | G | T |
| Overhang position | 0 | 1 | 2 | 3 | |

| | | | | | |
|-------------------|---------|---|---|---|---|
| SNP Y 10/14 | 5' GTTT | | | | |
| T allele | 3' CAAA | A | G | T | A |
| Overhang position | 1 | 2 | 3 | 4 | |

| | | | | | |
|-------------------|--------|---|---|---|---|
| SNP Y 11/15 | 5' GTT | | | | |
| T allele | 3' CAA | A | A | G | T |
| Overhang position | 0 | 1 | 2 | 3 | |

[0395] After fill-in with labeled ddATP and unlabeled dCTP, dGTP, and dTTP, the following molecules would be generated:

| | | | | | |
|-------------------|---------|----|---|---|---|
| SNP Y 10/14 | 5' GTTT | A* | | | |
| A allele | 3' CAAA | T | G | T | A |
| Overhang position | 1 | 2 | 3 | 4 | |

| | | | | | |
|-------------------|--------|---|----|---|---|
| SNP Y 11/15 | 5' GTT | T | A* | | |
| A allele | 3' CAA | A | T | G | T |
| Overhang position | 0 | 1 | 2 | 3 | |

| | Overhang position | 0 | 1 | 2 | 3 |
|-------------|-------------------|---|---|----|----|
| SNP Y 10/14 | 5' GTTT | T | C | A* | |
| T allele | 3' CAAA | A | G | T | A |
| | Overhang position | 1 | 2 | 3 | 4 |
| SNP Y 11/15 | 5' GTT | T | T | C | A* |
| T allele | 3' CAA | A | A | G | T |
| | Overhang position | 0 | 1 | 2 | 3 |

[0396] In this example, labeled ddGTP and labeled ddATP are used to determine the identity of both alleles of SNP X and SNP Y respectively. The nucleotide immediately preceding (the complementary nucleotide to position 0 of the overhang from the 11/15 cut SNP X is not guanine or adenine on the strand that is filled-in. Likewise, the nucleotide immediately preceding SNPY is not guanine or adenine on the strand that is filled-in. This allows the fill-in reaction for both SNPs to occur in a single reaction with labeled ddGTP, labeled ddATP, and unlabeled dCTP and dTTP. This reduces the number of reactions that need to be performed and increases the number of SNPs that can be analyzed in one reaction.

[0397] The first primers for each SNP can be designed to anneal at different distances from the locus of interest, which allows the SNPs to migrate at different positions on the gel. For example, the first primer used to amplify SNP X can anneal at 30 bases from the locus of interest, and the first primer used to amplify SNP Y can anneal at 35 bases from the locus of interest. Also, the nucleotides can be labeled with fluorescent dyes that emit at spectrums that do not overlap. After running the gel, the gel can be scanned at one wavelength specific for one dye. Only those molecules labeled with that dye will emit a signal. The gel then can be scanned at the wavelength for the second dye. Only those molecules labeled with

that dye will emit a signal. This method allows maximum compression for the number of SNPs that can be analyzed in a single reaction.

[0398] In this example, the nucleotide preceding the variable site on the strand that was filled-in is not be adenine or guanine. This method can work with any combination of labeled nucleotides, and the skilled artisan would understand which labeling reactions can be mixed and those that can not. For instance, if one SNP is labeled with thymidine and a second SNP is labeled with cytosine, the SNPs can be labeled in a single reaction if the nucleotide immediately preceding each variable site is not thymidine or cytosine on the sense strand and the nucleotide immediately after the variable site is not thymidine or cytosine on the sense strand.

[0399] This method allows the signals from one allele to be compared to the signal from a second allele without the added complexity of determining the degree of alternate cutting, or having to correct for the quantum coefficients of the dyes. This method is especially useful when trying to quantitate a ratio for one allele to another. For example, this method is useful for detecting chromosomal abnormalities. The ratio of alleles at a heterozygous site is expected to be about 1:1 (one A allele and one G allele). However, if an extra chromosome is present the ratio is expected to be about 1:2 (one A allele and 2 G alleles or 2 A alleles and 1 G allele). This method is especially useful when trying to detect fetal DNA in the presence of maternal DNA.

[0400] In addition, this method is useful for detecting two genetic signals in one sample. For example, this method can detect mutant cells in the presence of wild type cells (see Example 5). If a mutant cell contains a mutation in the DNA sequence of a particular gene, this method can be used to detect both the mutant signal and the wild type signal. This method can be used to detect the mutant DNA sequence in the presence of the wild type DNA sequence. The ratio of mutant DNA to wild type DNA can be quantitated because a single nucleotide labeled with one signal generating moiety is used.

EXAMPLE 5

[0401] Non-invasive methods for the detection of various types of cancer have the potential to reduce morbidity and mortality from the disease. Several techniques for the early detection of colorectal tumors have been developed including colonoscopy, barium enemas, and sigmoidoscopy but are limited in use because the techniques are invasive, which causes a low rate of patient compliance. Non-invasive genetic tests may be useful in identifying early stage colorectal tumors.

[0402] In 1991, researchers identified the Adenomatous Polyposis Coli gene (APC), which plays a critical role in the formation of colorectal tumors (Kinzler *et al.*, *Science* 253:661-665, 1991). The APC gene resides on chromosome 5q21-22 and a total of 15 exons code for an RNA molecule of 8529 nucleotides, which produces a 300 Kd APC protein. The protein is expressed in numerous cell types and is essential for cell adhesion.

[0403] Mutations in the APC gene generally initiate colorectal neoplasia (Tsao, J. *et al.*, *Am. J. Pathol.* 145:531-534, 1994). Approximately 95% of the mutations in the APC gene result in nonsense/frameshift mutations. The most common mutations occur at codons 1061 and 1309; mutations at these codons account for 1/3 of all germline mutations. With regard to somatic mutations, 60% occur within codons 1286-1513, which is about 10% of the coding sequence. This region is termed the mutation Cluster Region (MCR). Numerous types of mutations have been identified in the APC gene including nucleotide substitutions (see Table III), splicing errors (see Table IV), small deletions (see Table V), small insertions (see Table VI), small insertions/deletions (see Table VII), gross deletions (see Table VIII), gross insertions (see Table IX), and complex rearrangements (see Table X).

[0404] Researchers have attempted to identify cells harboring mutations in the APC gene in stool samples (Traverso, G. *et al.*, *New England Journal of Medicine*, Vol 346:311-320, 2002). While APC mutations are found in nearly all tumors, about 1 in 250 cells in the stool sample has a mutation in the APC gene;

most of the cells are normal cells that have been shed into the feces. Furthermore, human DNA represents about one-billionth of the total DNA found in stool samples; the majority of DNA is bacterial. The technique employed by Traverso *et al.* only detects mutations that result in a truncated protein.

[0405] As discussed above, numerous mutations in the APC gene have been implicated in the formation of colorectal tumors. Thus, there still exists a need for a highly sensitive, non-invasive technique for the detection of colorectal tumors. Below, methods are described for detection of two mutations in the APC gene. However, any number of mutations can be analyzed using the methods described herein.

Preparation of Template DNA

[0406] The template DNA is purified from a sample containing colon cells including but not limited to a stool sample. The template DNA is purified using the procedures described by Ahlquist *et al.* (Gastroenterology, 119:1219-1227, 2000). If stool samples are frozen, the samples are thawed at room temperature, and homogenized with an Exactor stool shaker (Exact Laboratories, Maynard, Mass.) Following homogenization, a 4 gram stool equivalent of each sample is centrifuged at 2536 x g for 5 minutes. The samples are centrifuged a second time at 16, 500 x g for 10 minutes. Supernatants are incubated with 20 μ l of RNase (0.5 mg per milliliter) for 1 hour at 37°C. DNA is precipitated with 1/10 volume of 3 mol of sodium acetate per liter and an equal volume of isopropanol. The DNA is dissolved in 5 ml of TRIS-EDTA (0.01 mol of Tris per liter (pH 7.4) and 0.001 mole of EDTA per liter.

Design of Primers

[0407] To determine if a mutation resides at codon 1370, the following primers are used:

First primer:

5' GTGCAAAGGCCTGAATTCCCAGGCACAAAGCTGTTGAA 3'

Second primer:

5' TGAAGCGAACTAGGGACTCAGGTGGACTT

[0408] The first primer contains a biotin tag at the extreme 5' end, and the nucleotide sequence for the restriction enzyme EcoRI. The second primer contains the nucleotide sequence for the restriction enzyme BsmF I.

[0409] To determine if a small deletion exists at codon 1302, the following primers are used:

First primer:

5' GATTCCGTAAACGAATTCAAGTTCAATTATCATCTTGTC 3'

Second primer:

5' CCATTGTTAACGCGGGACTTCTGCTATTG 3'

[0410] The first primer has a biotin tag at the 5' end and contains a restriction enzyme recognition site for EcoRI. The second primer contains a restriction enzyme recognition site for BsmF I.

PCR Reaction

[0411] The loci of interest are amplified from the template genomic DNA using the polymerase chain reaction (PCR, U.S. Patent Nos. 4,683,195 and 4,683,202, incorporated herein by reference). The loci of interest are amplified in separate reaction tubes; they can also be amplified together in a single PCR reaction. For increased specificity, a "hot-start" PCR reaction is used, e.g. by using the HotStarTaq Master Mix Kit supplied by QIAGEN (catalog number 203443). The amount of template DNA and primer per reaction are optimized for each locus of interest but in this example, 40 ng of template human genomic DNA and 5 μ M of each primer are used. Forty cycles of PCR are performed. The following PCR conditions are used:

- (1) 95°C for 15 minutes and 15 seconds;
- (2) 37°C for 30 seconds;
- (3) 95°C for 30 seconds;
- (4) 57°C for 30 seconds;
- (5) 95°C for 30 seconds;
- (6) 64°C for 30 seconds;

- (7) 95°C for 30 seconds;
- (8) Repeat steps 6 and 7 thirty nine (39) times;
- (9) 72°C for 5 minutes.

[0412] In the first cycle of PCR, the annealing temperature is about the melting temperature of the 3' annealing region of the second primers, which is 37°C. The annealing temperature in the second cycle of PCR is about the melting temperature of the 3' region, which anneals to the template DNA, of the first primer, which is 57°C. The annealing temperature in the third cycle of PCR is about the melting temperature of the entire sequence of the second primer, which is 64°C. The annealing temperature for the remaining cycles is 64°C. Escalating the annealing temperature from TM1 to TM2 to TM3 in the first three cycles of PCR greatly improves specificity. These annealing temperatures are representative, and the skilled artisan understands that the annealing temperatures for each cycle are dependent on the specific primers used.

[0413] The temperatures and times for denaturing, annealing, and extension, are optimized by trying various settings and using the parameters that yield the best results.

Purification of Fragment Containing Locus of Interest

[0414] The PCR products are separated from the genomic template DNA. Each PCR product is divided into four separate reaction wells of a Streptawell, transparent, High-Bind plate from Roche Diagnostics GmbH (catalog number 1 645 692, as listed in Roche Molecular Biochemicals, 2001 Biochemicals Catalog). The first primers contain a 5' biotin tag so the PCR products bound to the Streptavidin coated wells while the genomic template DNA does not. The streptavidin binding reaction is performed using a Thermomixer (Eppendorf) at 1000 rpm for 20 min. at 37°C. Each well is aspirated to remove unbound material, and washed three times with 1X PBS, with gentle mixing (Kandpal et al., Nucl. Acids Res. 18:1789-1795 (1990); Kaneoka et al., Biotechniques 10:30-34 (1991); Green et al., Nucl. Acids Res. 18:6163-6164 (1990)).

[0415] Alternatively, the PCR products are placed into a single well of a streptavidin plate to perform the nucleotide incorporation reaction in a single well.

Restriction Enzyme Digestion of Isolated Fragments Containing Loci of Interest

[0416] The purified PCR products are digested with the restriction enzyme BsmF I (New England Biolabs catalog number R0572S), which binds to the recognition site incorporated into the PCR products from the second primer. The digests are performed in the Streptawells following the instructions supplied with the restriction enzyme. After digestion with the appropriate restriction enzyme, the wells are washed three times with PBS to remove the cleaved fragments.

Incorporation of Labeled Nucleotide

[0417] The restriction enzyme digest described above yields a DNA fragment with a 5' overhang, which contains the locus of interest and a 3' recessed end. The 5' overhang functions as a template allowing incorporation of a nucleotide or nucleotides in the presence of a DNA polymerase.

[0418] For each locus of interest, four separate fill in reactions are performed; each of the four reactions contains a different fluorescently labeled ddNTP (ddATP, ddTTP, ddGTP, or ddCTP). The following components are added to each fill in reaction: 1 μ l of a fluorescently labeled ddNTP, 0.5 μ l of unlabeled ddNTPs (40 μ M), which contains all nucleotides except the nucleotide that is fluorescently labeled, 2 μ l of 10X sequenase buffer, 0.25 μ l of Sequenase, and water as needed for a 20 μ l reaction. The fill are performed in reactions at 40°C for 10 min. Non-fluorescently labeled ddNTP are purchased from Fermentas Inc. (Hanover, MD). All other labeling reagents are obtained from Amersham (Thermo Sequenase Dye Terminator Cycle Sequencing Core Kit, US 79565). In the presence of fluorescently labeled ddNTPs, the 3' recessed end is extended by one base, which corresponds to the locus of interest.

[0419] A mixture of labeled ddNTPs and unlabeled dNTPs also can be used for the fill-in reaction. The "fill in" conditions are as described above except that a mixture containing 40 μ M unlabeled dNTPs, 1 μ l fluorescently labeled ddATP,

1 μ l fluorescently labeled ddTTP, 1 μ l fluorescently labeled ddCTP, and 1 μ l ddGTP are used. The fluorescent ddNTPs are obtained from Amersham (Thermo Sequenase Dye Terminator Cycle Sequencing Core Kit, US 79565; Amersham does not publish the concentrations of the fluorescent nucleotides). The locus of interest is digested with the restriction enzyme BsmF I, which generates a 5' overhang of four bases. If the first nucleotide incorporated is a labeled ddNTP, the 3' recessed end is filled in by one base, allowing detection of the locus of interest. However, if the first nucleotide incorporated is a dNTP, the polymerase continues to incorporate nucleotides until a ddNTP is filled in. For example, the first two nucleotides may be filled in with dNTPs, and the third nucleotide with a ddNTP, allowing detection of the third nucleotide in the overhang. Thus, the sequence of the entire 5' overhang is determined, which increases the information obtained from each SNP or locus of interest. This type of fill in reaction is especially useful when detecting the presence of insertions, deletions, insertions and deletions, rearrangements, and translocations.

[0420] Alternatively, one nucleotide labeled with a single dye is used to determine the sequence of the locus of interest. See Example 4. This method eliminates any potential errors when using different dyes, which have different quantum coefficients.

[0421] After labeling, each Streptawell is rinsed with 1X PBS (100 μ l) three times. The "filled in" DNA fragments are released from the Streptawells by digesting with the restriction enzyme EcoRI, according to the manufacturer's instructions that are supplied with the enzyme. The digestion is performed for 1 hour at 37 °C with shaking at 120 rpm.

Detection of the Locus of Interest

[0422] After release from the streptavidin matrix, the sample is loaded into a lane of a 36 cm 5% acrylamide (urea) gel (BioWhittaker Molecular Applications, Long Ranger Run Gel Packs, catalog number 50691). The sample is electrophoresed into the gel at 3000 volts for 3 min. The gel is run for 3 hours

using a sequencing apparatus (Hoefer SQ3 Sequencer). The incorporated labeled nucleotide is detected by fluorescence.

[0423] To determine if any cells contain mutations at codon 1370 of the APC gene when separate fill-in reactions are performed, the lanes of the gel that correspond to the fill-in reaction for ddATP and ddTTP are analyzed. If only normal cells are present, the lane corresponding to the fill in reaction with ddATP is a bright signal. No signal is detected for the "fill-in" reaction with ddTTP. However, if the patient sample contains cells with mutations at codon 1370 of the APC gene, the lane corresponding to the fill in reaction with ddATP is a bright signal, and a signal is detected from the lane corresponding to the fill in reaction with ddTTP. The intensity of the signal from the lane corresponding to the fill in reaction with ddTTP is indicative of the number of mutant cells in the sample.

[0424] Alternatively, one labeled nucleotide is used to determine the sequence of the alleles at codon 1370 of the APC gene. At codon 1370, the normal sequence is AAA, which codes for the amino acid lysine. However, a nucleotide substitution has been identified at codon 1370, which is associated with colorectal tumors. Specifically, a change from A to T (AAA-TAA) typically is found at codon 1370, which results in a stop codon. A single fill-in reaction is performed using labeled ddATP, and unlabeled dTTP, dCTP, and dGTP. A single nucleotide labeled with one fluorescent dye is used to determine the presence of both the normal and mutant DNA sequence that codes for codon 1370. The relevant DNA sequence is depicted below with the sequence corresponding to codon 1370 in bold:

5' CCCAAAAGTCCACCTGA
3' GGGTTTCAGGTGGACT

[0425] After digest with BsmF I, the following overhang is produced:

| | | | | | |
|-------------------|--------|---|---|---|---|
| | 5' CCC | | | | |
| | 3' GGG | T | T | T | T |
| Overhang position | | 1 | 2 | 3 | 4 |

[0426] If the patient sample has no cells harboring a mutation at codon 1370, one signal is seen corresponding to incorporation of labeled ddATP.

| | | | | | |
|-------------------|--------|----|---|---|---|
| | 5' CCC | A* | | | |
| | 3' GGG | T | T | T | T |
| Overhang position | | 1 | 2 | 3 | 4 |

[0427] However, if the patient sample has cells with mutations at codon 1370 of the APC gene, one signal is seen, which corresponds to the normal sequence at codon 1370, and a second signal is seen, which corresponds to the mutant sequence at codon 1370. The signals clearly are identified as they differ in molecular weight.

| | | | | | |
|----------------------------------|-----|---|---|---|---|
| Overhang of normal DNA sequence: | CCC | | | | |
| | GGG | T | T | T | T |
| Overhang position | | 1 | 2 | 3 | 4 |

| | | | | | |
|------------------------------------|-----|----|---|---|---|
| Normal DNA sequence after fill-in: | CCC | A* | | | |
| | GGG | T | T | T | T |
| Overhang position | | 1 | 2 | 3 | 4 |

| | | | | | |
|----------------------------------|-----|---|---|---|---|
| Overhang of mutant DNA sequence: | CCC | | | | |
| | GGG | A | T | T | T |
| Overhang position | | 1 | 2 | 3 | 4 |

| | | | | | |
|------------------------------------|-----|---|----|---|---|
| Mutant DNA sequence after fill-in: | CCC | T | A* | | |
| | GGG | A | T | T | T |
| Overhang position | | 1 | 2 | 3 | 4 |

[0428] Two signals are seen when the mutant allele is present. The mutant DNA molecules are filled in one base after the wild type DNA molecules. The two signals are separated using any method that discriminates based on molecular

weight. One labeled nucleotide (ddATP) is used to detect the presence of both the wild type DNA sequence and the mutant DNA sequence. This method of labeling reduces the number of reactions that need to be performed and allows accurate quantitation for the number of mutant cells in the patient sample. The number of mutant cells in the sample is used to determine patient prognosis, the degree and the severity of the disease. This method of labeling eliminates the complications associated with using different dyes, which have distinct quantum coefficients. This method of labeling also eliminates errors associated with pipetting reactions.

[0429] To determine if any cells contain mutations at codon 1302 of the APC gene when separate fill-in reactions are performed, the lanes of the gel that correspond to the fill-in reaction for ddTTP and ddCTP are analyzed. The normal DNA sequence is depicted below with sequence coding for codon 1302 in bold type-face.

Normal Sequence: 5' ACCCT**G**CAAATAGCAGAA
 3' TGGGACGTT TATCGTCT T

[0430] After digest, the following 5' overhang is produced:

| | | | | | |
|-------------------|---------|---|---|---|---|
| | 5' ACCC | | | | |
| | 3' TGGG | A | C | G | T |
| Overhang position | | 1 | 2 | 3 | 4 |

[0431] After the fill-in reaction, labeled ddTTP is incorporated.

| | | | | | |
|-------------------|---------|----|---|---|---|
| | 5' ACCC | T* | | | |
| | 3' TGGG | A | C | G | T |
| Overhang position | | 1 | 2 | 3 | 4 |

[0432] A deletion of a single base of the APC sequence, which typically codes for codon 1302, has been associated with colorectal tumors. The mutant DNA sequence is depicted below with the relevant sequence in bold:

Mutant Sequence: 5' ACCCGCAAATAGCAGAA
 3' TGGGCGTTATCGTCCTT

After digest:

| | | | | | |
|-------------------|--------|---|---|---|---|
| | 5' ACC | | | | |
| | 3' TGG | G | C | G | T |
| Overhang position | | 1 | 2 | 3 | 4 |

After fill-in:

| | | | | | |
|-------------------|--------|----|---|---|---|
| | 5' ACC | C* | | | |
| | 3' TGG | G | C | G | T |
| Overhang position | | 1 | 2 | 3 | 4 |

[0433] If there are no mutations in the APC gene, signal is not detected for the fill in reaction with ddCTP*, but a bright signal is detected for the fill-in reaction with ddTTP*. However, if there are cells in the patient sample that have mutations in the APC gene, signals are seen for the fill-in reactions with ddCTP* and ddTTP*.

[0434] Alternatively, a single fill-in reaction is performed using a mixture containing unlabeled dNTPs, fluorescently labeled ddATP, fluorescently labeled ddTTP, fluorescently labeled ddCTP, and fluorescently labeled ddGTP. If there is no deletion, labeled ddTTP is incorporated.

| | | | | | |
|-------------------|---------|----|---|---|---|
| | 5' ACCC | T* | | | |
| | 3' TGGG | A | C | G | T |
| Overhang position | | 1 | 2 | 3 | 4 |

[0435] However, if the T has been deleted, labeled ddCTP* is incorporated.

| | | | | | |
|-------------------|--------|----|---|---|---|
| | 5' ACC | C* | | | |
| | 3' TGG | G | C | G | T |
| Overhang position | | 1 | 2 | 3 | 4 |

[0436] The two signals are separated by molecular weight because of the deletion of the thymidine nucleotide. If mutant cells are present, two signals are generated in the same lane but are separated by a single base pair (this principle is demonstrated in FIG 9D). The deletion causes a change in the molecular weight of the DNA fragments, which allows a single fill in reaction to be used to detect the presence of both normal and mutant cells.

[0437] In the above example, methods for the detection of a nucleotide substitution and a small deletion are described. However, the methods are used for the detection of any type of mutation including but not limited to nucleotide substitutions (see Table III), splicing errors (see Table IV), small deletions (see Table V), small insertions (see Table VI), small insertions/deletions (see Table VII), gross deletions (see Table VIII), gross insertions (see Table IX), and complex rearrangements (see Table X).

[0438] In addition, the above-described methods are used for the detection of any type of disease including but not limited to those listed in Table II. Furthermore, any type of mutant gene is detected using the inventions described herein including but not limited to the genes associated with the diseases listed in Table II, BRCA1, BRCA2, MSH6, MSH2, MLH1, RET, PTEN, ATM, H-RAS, p53, ELAC2, CDH1, APC, AR, PMS2, MLH3, CYP1A1, GSTP1, GSTM1, AXIN2, CYP19, MET, NAT1, CDKN2A, NQO1, trc8, RAD51, PMS1, TGFBR2, VHL, MC4R, POMC, NROB2, UCP2, PCSK1, PPARG, ADRB2, UCP3, glur1, cart, SORBS1, LEP, LEPR, SIM1, TNF, IL-6, IL-1, IL-2, IL-3, IL1A, TAP2, THPO, THRB, NBS1, RBM15, LIF, MPL, RUNX1, Her-2, glucocorticoid receptor, estrogen receptor, thyroid receptor, p21, p27, K-RAS, N-RAS,

retinoblastoma protein, Wiskott-Aldrich (WAS) gene, Factor V Leiden, Factor II (prothrombin), methylene tetrahydrofolate reductase, cystic fibrosis, LDL receptor, HDL receptor, superoxide dismutase gene, SHOX gene, genes involved in nitric oxide regulation, genes involved in cell cycle regulation, tumor suppressor genes, oncogenes, genes associated with neurodegeneration, genes associated with obesity, . Abbreviations correspond to the proteins as listed on the Human Gene Mutation Database, which is incorporated herein by reference (www.archive.uwcm.ac.uk/uwcm) website address active as of February 12, 2003).

[0439] The above-example demonstrates the detection of mutant cells and mutant alleles from a fecal sample. However, the methods described herein are used for detection of mutant cells from any biological sample including but not limited to blood sample, serum sample, plasma sample, urine sample, spinal fluid, lymphatic fluid, semen, vaginal secretion, ascitic fluid, saliva, mucosa secretion, peritoneal fluid, fecal sample, body exudates, breast fluid, lung aspirates, cells, tissues, individual cells or extracts of the such sources that contain the nucleic acid of the same, and subcellular structures such as mitochondria or chloroplasts. In addition, the methods described herein are used for the detection of mutant cells and mutated DNA from any number of nucleic acid containing sources including but not limited to forensic, food, archeological, agricultural or inorganic samples.

[0440] The above example is directed to detection of mutations in the APC gene. However, the inventions described herein are used for the detection of mutations in any gene that is associated with or predisposes to disease (see Table XI).

[0441] For example, hypermethylation of the glutathione S-transferase P1 (GSTP1) promoter is the most common DNA alteration in prostate cancer. The methylation state of the promoter is determined using sodium bisulfite and the methods described herein.

[0442] Treatment with sodium bisulfite converts unmethylated cytosine residues into uracil, and leaving the methylated cytosines unchanged. Using the methods described herein, a first and second primer are designed to amplify the regions of the GSTP1 promoter that are often methylated. Below, a region of the GSTP1 promoter is shown prior to sodium bisulfite treatment:

[0443] Before Sodium Bisulfite treatment:

5' ACCGCTACA

3' TGGCGATCA

[0444] Below, a region of the GSTP1 promoter is shown after sodium bisulfite treatment, PCR amplification, and digestion with the type IIS restriction enzyme BsmF I:

Unmethylated

| | | | | |
|-------------------|---|---|---|---|
| 5' ACC | | | | |
| 3' TGG | U | G | A | T |
| Overhang position | 1 | 2 | 3 | 4 |

Methylated

| | | | | |
|-------------------|---|---|---|---|
| 5' ACC | | | | |
| 3' TGG | C | G | A | T |
| Overhang position | 1 | 2 | 3 | 4 |

[0445] Labeled ddATP, unlabeled dCTP, dGTP, and dTTP are used to fill-in the 5' overhangs. The following molecules are generated:

Unmethylated

| | | | | |
|-------------------|---|----|---|---|
| 5' ACC | | A* | | |
| 3' TGG | U | G | A | T |
| Overhang position | 1 | 2 | 3 | 4 |

Methylated

| | | | | |
|-------------------|---|---|---|----|
| 5' ACC | G | C | T | A* |
| 3' TGG | C | G | A | T |
| Overhang position | 1 | 2 | 3 | 4 |

[0446] Two signals are seen; one corresponds to DNA molecules filled in with ddATP at position one complementary to the overhang (unmethylated), and the other corresponds to the DNA molecules filled in with ddATP at position 4 complementary to the overhang (methylated). The two signals are separated based on molecular weight. Alternatively, the fill-in reactions are performed in separate reactions using labeled ddGTP in one reaction and labeled ddATP in another reaction.

[0447] The methods described herein are used to screen for prostate cancer and also to monitor the progression and severity of the disease. The use of a single nucleotide to detect both the methylated and unmethylated sequences allows accurate quantitation and provides a high level of sensitivity for the methylated sequences, which is a useful tool for earlier detection of the disease.

[0448] The information contained in Tables III-X was obtained from the Human Gene Mutation Database. With the information provided herein, the skilled artisan will understand how to apply these methods for determining the sequence of the alleles for any gene. A large number of genes and their associated mutations can be found at the following website:

www.archive.uwcm.ac.uk./uwcm.

TABLE III:
NUCLEOTIDE SUBSTITUTIONS

| Codon | Nucleotide | Amino acid | Phenotype |
|-------|------------|------------|----------------------------|
| 99 | CGG-TGG | Arg-Trp | Adenomatous polyposis coli |
| 121 | AGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 157 | TGG-TAG | Trp-Term | Adenomatous polyposis coli |
| 159 | TAC-TAG | Tyr-Term | Adenomatous polyposis coli |
| 163 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 168 | AGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 171 | AGT-ATT | Ser-Ile | Adenomatous polyposis coli |
| 181 | CAA-TAA | Gln-Term | Adenomatous polyposis coli |
| 190 | GAA-TAA | Glu-Term | Adenomatous polyposis coli |

| | | | |
|-----|---------|----------|----------------------------|
| 202 | GAA-TAA | Glu-Term | Adenomatous polyposis coli |
| 208 | CAG-CGG | Gln-Arg | Adenomatous polyposis coli |
| 208 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 213 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 215 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 216 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 232 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 233 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 247 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 267 | GGA-TGA | Gly-Term | Adenomatous polyposis coli |
| 278 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 280 | TCA-TGA | Ser-Term | Adenomatous polyposis coli |
| 280 | TCA-TAA | Ser-Term | Adenomatous polyposis coli |
| 283 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 302 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 332 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 358 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 405 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 414 | CGC-TGC | Arg-Cys | Adenomatous polyposis coli |
| 422 | GAG-TAG | Glu-Term | Adenomatous polyposis coli |
| 423 | TGG-TAG | Trp-Term | Adenomatous polyposis coli |
| 424 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 433 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 443 | GAA-TAA | Glu-Term | Adenomatous polyposis coli |
| 457 | TCA-TAA | Ser-Term | Adenomatous polyposis coli |
| 473 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 486 | TAC-TAG | Tyr-Term | Adenomatous polyposis coli |
| 499 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 500 | TAT-TAG | Tyr-Term | Adenomatous polyposis coli |
| 541 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 553 | TGG-TAG | Trp-Term | Adenomatous polyposis coli |
| 554 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 564 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 577 | TTA-TAA | Leu-Term | Adenomatous polyposis coli |
| 586 | AAA-TAA | Lys-Term | Adenomatous polyposis coli |
| 592 | TTA-TGA | Leu-Term | Adenomatous polyposis coli |
| 593 | TGG-TAG | Trp-Term | Adenomatous polyposis coli |
| 593 | TGG-TGA | Trp-Term | Adenomatous polyposis coli |
| 622 | TAC-TAA | Tyr-Term | Adenomatous polyposis coli |
| 625 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 629 | TTA-TAA | Leu-Term | Adenomatous polyposis coli |
| 650 | GAG-TAG | Glu-Term | Adenomatous polyposis coli |
| 684 | TTG-TAG | Leu-Term | Adenomatous polyposis coli |
| 685 | TGG-TGA | Trp-Term | Adenomatous polyposis coli |

| | | | |
|------|---------|----------|----------------------------|
| 695 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 699 | TGG-TGA | Trp-Term | Adenomatous polyposis coli |
| 699 | TGG-TAG | Trp-Term | Adenomatous polyposis coli |
| 713 | TCA-TGA | Ser-Term | Adenomatous polyposis coli |
| 722 | AGT-GGT | Ser-Gly | Adenomatous polyposis coli |
| 747 | TCA-TGA | Ser-Term | Adenomatous polyposis coli |
| 764 | TTA-TAA | Leu-Term | Adenomatous polyposis coli |
| 784 | TCT-ACT | Ser-Thr | Adenomatous polyposis coli |
| 805 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 811 | TCA-TGA | Ser-Term | Adenomatous polyposis coli |
| 848 | AAA-TAA | Lys-Term | Adenomatous polyposis coli |
| 876 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 879 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 893 | GAA-TAA | Glu-Term | Adenomatous polyposis coli |
| 932 | TCA-TAA | Ser-Term | Adenomatous polyposis coli |
| 932 | TCA-TGA | Ser-Term | Adenomatous polyposis coli |
| 935 | TAC-TAG | Tyr-Term | Adenomatous polyposis coli |
| 935 | TAC-TAA | Tyr-Term | Adenomatous polyposis coli |
| 995 | TGC-TGA | Cys-Term | Adenomatous polyposis coli |
| 997 | TAT-TAG | Tyr-Term | Adenomatous polyposis coli |
| 999 | CAA-TAA | Gln-Term | Adenomatous polyposis coli |
| 1000 | TAC-TAA | Tyr-Term | Adenomatous polyposis coli |
| 1020 | GAA-TAA | Glu-Term | Adenomatous polyposis coli |
| 1032 | TCA-TAA | Ser-Term | Adenomatous polyposis coli |
| 1041 | CAA-TAA | Gln-Term | Adenomatous polyposis coli |
| 1044 | TCA-TAA | Ser-Term | Adenomatous polyposis coli |
| 1045 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 1049 | TGG-TGA | Trp-Term | Adenomatous polyposis coli |
| 1067 | CAA-TAA | Gln-Term | Adenomatous polyposis coli |
| 1071 | CAA-TAA | Gln-Term | Adenomatous polyposis coli |
| 1075 | TAT-TAA | Tyr-Term | Adenomatous polyposis coli |
| 1075 | TAT-TAG | Tyr-Term | Adenomatous polyposis coli |
| 1102 | TAC-TAG | Tyr-Term | Adenomatous polyposis coli |
| 1110 | TCA-TGA | Ser-Term | Adenomatous polyposis coli |
| 1114 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 1123 | CAA-TAA | Gln-Term | Adenomatous polyposis coli |
| 1135 | TAT-TAG | Tyr-Term | Adenomatous polyposis coli |
| 1152 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 1155 | GAA-TAA | Glu-Term | Adenomatous polyposis coli |
| 1168 | GAA-TAA | Glu-Term | Adenomatous polyposis coli |
| 1175 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 1176 | CCT-CTT | Pro-Leu | Adenomatous polyposis coli |
| 1184 | GCC-CCC | Ala-Pro | Adenomatous polyposis coli |
| 1193 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |

| | | | |
|------|---------|----------|---|
| 1194 | TCA-TGA | Ser-Term | Adenomatous polyposis coli |
| 1198 | TCA-TGA | Ser-Term | Adenomatous polyposis coli |
| 1201 | TCA-TGA | Ser-Term | Adenomatous polyposis coli |
| 1228 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 1230 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 1244 | CAA-TAA | Gln-Term | Adenomatous polyposis coli |
| 1249 | TGC-TGA | Cys-Term | Adenomatous polyposis coli |
| 1256 | CAA-TAA | Gln-Term | Adenomatous polyposis coli |
| 1262 | TAT-TAA | Tyr-Term | Adenomatous polyposis coli |
| 1270 | TGT-TGA | Cys-Term | Adenomatous polyposis coli |
| 1276 | TCA-TGA | Ser-Term | Adenomatous polyposis coli |
| 1278 | TCA-TAA | Ser-Term | Adenomatous polyposis coli |
| 1286 | GAA-TAA | Glu-Term | Adenomatous polyposis coli |
| 1289 | TGT-TGA | Cys-Term | Adenomatous polyposis coli |
| 1294 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 1307 | ATA-AAA | Ile-Lys | Colorectal cancer, predisposition to, association |
| 1309 | GAA-TAA | Glu-Term | Adenomatous polyposis coli |
| 1317 | GAA-CAA | Glu-Gln | Colorectal cancer, predisposition to |
| 1328 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 1338 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 1342 | TTA-TAA | Leu-Term | Adenomatous polyposis coli |
| 1342 | TTA-TGA | Leu-Term | Adenomatous polyposis coli |
| 1348 | AGG-TGG | Arg-Trp | Adenomatous polyposis coli |
| 1357 | GGA-TGA | Gly-Term | Adenomatous polyposis coli |
| 1367 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 1370 | AAA-TAA | Lys-Term | Adenomatous polyposis coli |
| 1392 | TCA-TAA | Ser-Term | Adenomatous polyposis coli |
| 1392 | TCA-TGA | Ser-Term | Adenomatous polyposis coli |
| 1397 | GAG-TAG | Glu-Term | Adenomatous polyposis coli |
| 1449 | AAG-TAG | Lys-Term | Adenomatous polyposis coli |
| 1450 | CGA-TGA | Arg-Term | Adenomatous polyposis coli |
| 1451 | GAA-TAA | Glu-Term | Adenomatous polyposis coli |
| 1503 | TCA-TAA | Ser-Term | Adenomatous polyposis coli |
| 1517 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 1529 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 1539 | TCA-TAA | Ser-Term | Adenomatous polyposis coli |
| 1541 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |
| 1564 | TTA-TAA | Leu-Term | Adenomatous polyposis coli |
| 1567 | TCA-TGA | Ser-Term | Adenomatous polyposis coli |
| 1640 | CGG-TGG | Arg-Trp | Adenomatous polyposis coli |
| 1693 | GAA-TAA | Glu-Term | Adenomatous polyposis coli |
| 1822 | GAC-GTC | Asp-Val | Adenomatous polyposis coli, association with ? |
| 2038 | CTG-GTG | Leu-Val | Adenomatous polyposis coli |
| 2040 | CAG-TAG | Gln-Term | Adenomatous polyposis coli |

| | | | |
|------|---------|---------|----------------------------|
| 2566 | AGA-AAA | Arg-Lys | Adenomatous polyposis coli |
| 2621 | TCT-TGT | Ser-Cys | Adenomatous polyposis coli |
| 2839 | CTT-TTT | Leu-Phe | Adenomatous polyposis coli |

TABLE IV:
NUCLEOTIDE SUBSTITUTIONS

| Donor/ Acceptor | Relative location | Substitution | Phenotype |
|--------------------|----------------------|--------------|----------------------------|
| ds | -1 | G-C | Adenomatous polyposis coli |
| as | -1 | G-A | Adenomatous polyposis coli |
| as | -1 | G-C | Adenomatous polyposis coli |
| ds | +2 | T-A | Adenomatous polyposis coli |
| as | -1 | G-C | Adenomatous polyposis coli |
| as | -1 | G-T | Adenomatous polyposis coli |
| as | -1 | G-A | Adenomatous polyposis coli |
| as | -2 | A-C | Adenomatous polyposis coli |
| as | -5 | A-G | Adenomatous polyposis coli |
| ds | +3 | A-C | Adenomatous polyposis coli |
| as | -1 | G-A | Adenomatous polyposis coli |
| ds | +1 | G-A | Adenomatous polyposis coli |
| as | -1 | G-T | Adenomatous polyposis coli |
| ds | +1 | G-A | Adenomatous polyposis coli |
| as | -1 | G-A | Adenomatous polyposis coli |
| ds | +1 | G-A | Adenomatous polyposis coli |
| ds | +3 | A-G | Adenomatous polyposis coli |
| ds | +5 | G-T | Adenomatous polyposis coli |
| as | -1 | G-A | Adenomatous polyposis coli |
| as | -6 | A-G | Adenomatous polyposis coli |
| as | -5 | A-G | Adenomatous polyposis coli |
| as | -2 | A-G | Adenomatous polyposis coli |
| ds | +2 | T-C | Adenomatous polyposis coli |
| as | -2 | A-G | Adenomatous polyposis coli |
| ds | +1 | G-A | Adenomatous polyposis coli |
| ds | +1 | G-T | Adenomatous polyposis coli |
| ds | +2 | T-G | Adenomatous polyposis coli |

TABLE V:
APC SMALL DELETIONS

[0449] Bold letters indicate the codon. Undercase letters represent the deletion. Where deletions extend beyond the coding region, other positional

information is provided. For example, the abbreviation 5' UTR represents 5' untranslated region, and the abbreviation E6I6 denotes exon 6/intron 6 boundary.

| Location/ codon | Deletion | Phenotype |
|--------------------|-------------------------------|----------------------------|
| 77 | TTAgataGCAGTAATT | Adenomatous polyposis coli |
| 97 | GGAAAGccgggaagGATCTGTATC | Adenomatous polyposis coli |
| 138 | GAGAaAGAGAG_E3I3_GTAA | Adenomatous polyposis coli |
| 139 | AAAGAgag_E3I3_Gtaactttct | Thyroid cancer |
| 139 | AAAAGagag_E3I3_GTAACCTTTC | Adenomatous polyposis coli |
| 142 | TTTTAAAAAAaAAAAATAG_I3E4_GTCA | Adenomatous polyposis coli |
| 144 | AAAATAG_I3E4_GTCatTGCTTCTTGC | Adenomatous polyposis coli |
| 149 | GACAaaGAAGAAAAGG | Adenomatous polyposis coli |
| 149 | GACAAagaaGAAAAGGAAA | Adenomatous polyposis coli |
| 155 | AGGAA^AAAGActggATTACGCTCA | Adenomatous polyposis coli |
| 169 | AAAAGA^ATAGatagTCTCCTTTA | Adenomatous polyposis coli |
| 172 | AGATAGT^CTTcCTTAACTGA | Adenomatous polyposis coli |
| 179 | TCCITTacaaACAGATATGA | Adenomatous polyposis coli |
| 185 | ACCaGAAGGCAATT | Adenomatous polyposis coli |
| 196 | ATCAGagTTGCGATGGA | Adenomatous polyposis coli |
| 213 | CGAGCaCAG_E5I5_GTAAGTT | Adenomatous polyposis coli |
| 298 | CAcTcTGCACCTCGA | Adenomatous polyposis coli |
| 329 | GATaTGTGCGCGAAC | Adenomatous polyposis coli |
| 365 | AAAGActCTGTATTGTT | Adenomatous polyposis coli |
| 397 | GACaaGAGAGGCAGG | Adenomatous polyposis coli |
| 427 | CATGAacCAGGCATGGA | Adenomatous |

| | | |
|-----|-------------------------------|-------------------------------|
| | | polyposis coli |
| 428 | GAACCaGGCATGGACC | Adenomatous polyposis coli |
| 436 | AATCCaa_E9I9_gTATGTTCTCT | Adenomatous polyposis coli |
| 440 | GCTCCtGTTGAACATC | Adenomatous polyposis coli |
| 455 | AAACTtTCATTGATG | Adenomatous polyposis coli |
| 455 | AAACttcaTTTGATGAAG | Adenomatous polyposis coli |
| 472 | CTAcAGGCCATTGC | Adenomatous polyposis coli |
| 472 | TAAATTAG_I10E11_GGgGACTACAGGC | Adenomatous polyposis coli |
| 478 | TTATtGCAAGTGGAC | Adenomatous polyposis coli |
| 486 | TACGgGCTTACTAAT | Adenomatous polyposis coli |
| 494 | AGTATtACACTAAGAC | Adenomatous polyposis coli |
| 495 | ATTACAcTAAGACGATA | Adenomatous polyposis coli |
| 497 | CTAaGACGATATGC | Adenomatous polyposis coli |
| 520 | TGCTCtaTGAAAGGCTG | Adenomatous polyposis coli |
| 526 | ATGAGagcatttgGCCCACTAA | Adenomatous polyposis coli |
| 539 | GACTTaCAGCAG_E12I12_GTAC | Adenomatous polyposis coli |
| 560 | AAAAAgaCGTTGCGAGA | Adenomatous polyposis coli |
| 566 | GTTGgaagtGTGAAAGCAT | Adenomatous polyposis coli |
| 570 | AAAGCaTTGATGGAAT | Adenomatous polyposis coli |
| 577 | TTAGaagtTAAAAAG_E13I13_GTA | Adenomatous polyposis coli |
| 584 | ACCCtCAAAAGCGTAT | Adenomatous polyposis coli |
| 591 | GCCTtATGGAATTG | Adenomatous polyposis coli |
| 608 | GCTgTAGATGGTGC | Adenomatous |

| | | |
|-----|--|-------------------------------|
| | | polyposis coli |
| 617 | GTTggactcttacttaccGGAGCCAGAC | Adenomatous polyposis coli |
| 620 | CTTACttacCGGAGCCAGA | Adenomatous polyposis coli |
| 621 | ACTTaCCGGAGCCAG | Adenomatous polyposis coli |
| 624 | AGCcaGACAAACACT | Adenomatous polyposis coli |
| 624 | AGCCagacAAACACTTTA | Adenomatous polyposis coli |
| 626 | ACAAacaCTTAGCCAT | Adenomatous polyposis coli |
| 629 | TTAGCcATTATTGAAA | Adenomatous polyposis coli |
| 635 | GGAGgTGGGATATTA | Adenomatous polyposis coli |
| 638 | ATATtACGGAATGTG | Adenomatous polyposis coli |
| 639 | TTACGgAATGTGTCCA | Adenomatous polyposis coli |
| 657 | AGAgaGAACAACTGT | Adenomatous polyposis coli |
| 659 | TATTCAG_I14E15_GCaaatcctaagagagAACA ACTGTC | Adenomatous polyposis coli |
| 660 | AACTgtCTACAAACTT | Adenomatous polyposis coli |
| 665 | TTAttACAACACTTA | Adenomatous polyposis coli |
| 668 | CACttAAAATCTCAT | Adenomatous polyposis coli |
| 673 | AGTttgacaatagtCAGTAATGCA | Adenomatous polyposis coli |
| 768 | CACTTaTCAGAAACTT | Adenomatous polyposis coli |
| 769 | TTATcAGAAACTTT | Adenomatous polyposis coli |
| 770 | TCAGAaACTTTGACA | Adenomatous polyposis coli |
| 780 | AGTCcCAAGGCATCT | Adenomatous polyposis coli |
| 792 | AAGCaAAGTCTCTAT | Adenomatous polyposis coli |
| 792 | AAGCAaaGTCTCTATGG | Adenomatous |

| | | |
|-----|-----------------------------|-------------------------------|
| | | polyposis coli |
| 793 | CAAAgTCTCTATGGT | Adenomatous polyposis coli |
| 798 | GATTatGTTTTGACA | Adenomatous polyposis coli |
| 802 | GACACcaatcgacatGATGATAATA | Adenomatous polyposis coli |
| 805 | CGACatGATGATAATA | Adenomatous polyposis coli |
| 811 | TCAGacaaTTTTAATACT | Adenomatous polyposis coli |
| 825 | TATTGAATACTAC | Adenomatous polyposis coli |
| 827 | AATAcTACAGTGTAA | Adenomatous polyposis coli |
| 830 | GTGTTacccagtcctctTCATCAAGAG | Adenomatous polyposis coli |
| 833 | AGCTCcTCTTCATCAA | Adenomatous polyposis coli |
| 836 | TCATcAAGAGGAAGC | Adenomatous polyposis coli |
| 848 | AAAGAtaGAAGTTGGAA | Adenomatous polyposis coli |
| 848 | AAAAGatagaagTTTGGAGAGA | Adenomatous polyposis coli |
| 855 | GAACgCGGAATTGGT | Adenomatous polyposis coli |
| 856 | CGCGgaattGGTCTAGGCA | Adenomatous polyposis coli |
| 856 | CGCGgAATTGGTCTA | Adenomatous polyposis coli |
| 879 | CAGaTCTCCACCAAC | Adenomatous polyposis coli |
| 902 | GAAGAcagaAGTTCTGGGT | Adenomatous polyposis coli |
| 907 | GGGTcTACCACTGAA | Adenomatous polyposis coli |
| 915 | GTGACaGATGAGAGAA | Adenomatous polyposis coli |
| 929 | CATACacatTCAAACACTT | Adenomatous polyposis coli |
| 930 | ACACAttcaAACACTTACA | Adenomatous polyposis coli |
| 931 | CATtCAAACACTTA | Adenomatous |

| | | |
|------|-------------------------------|----------------------------|
| | | polyposis coli |
| 931 | CATTcAAACACTTAC | Adenomatous polyposis coli |
| 933 | AACacttACAATTCAC | Adenomatous polyposis coli |
| 935 | TACAAattcactAAGTCGGAAA | Adenomatous polyposis coli |
| 937 | TTCActaaGTCGGAAAAT | Adenomatous polyposis coli |
| 939 | AAGtcggAAAATTCAAA | Adenomatous polyposis coli |
| 946 | ACATgTTCTATGCCT | Adenomatous polyposis coli |
| 954 | TTAGaaTACAAGAGAT | Adenomatous polyposis coli |
| 961 | AATgATAGTTAAA | Adenomatous polyposis coli |
| 963 | AGTTTaAATAGTGTCA | Adenomatous polyposis coli |
| 964 | TTAaataGTGTCAGTAG | Adenomatous polyposis coli |
| 973 | TATGgTAAAAGAGGT | Adenomatous polyposis coli |
| 974 | GGTAAaAGAGGTCAAA | Adenomatous polyposis coli |
| 975 | AAAAGaGGTCAAATGA | Thyroid cancer |
| 992 | AGTAAGTTTGCAGTT | Thyroid cancer |
| 993 | AAGtttgcatTAAGGTCAATAC | Adenomatous polyposis coli |
| 999 | CAAtacccagCCGACCTAGC | Adenomatous polyposis coli |
| 1023 | ACACcAATAAATTAT | Adenomatous polyposis coli |
| 1030 | AAAtATTCAAGATGA | Adenomatous polyposis coli |
| 1032 | TCAGatgagCAGTTGAAC | Adenomatous polyposis coli |
| 1033 | GATGaGCAGTTGAAC | Adenomatous polyposis coli |
| 1049 | TGGGcAAGACCCAAA | Adenomatous polyposis coli |
| 1054 | CACAtaataGAAGATGAAA | Adenomatous polyposis coli |
| 1055 | ATAAtagaaGATGAAATAA | Adenomatous |

| | | |
|------|-----------------------------|----------------------------|
| | | polyposis coli |
| 1056 | ATAGAaGATGAAATAA | Adenomatous polyposis coli |
| 1060 | ATAAAacaaaGTGAGCAAAG | Adenomatous polyposis coli |
| 1061 | AAAacaaaGTGAGCAAAG | Adenomatous polyposis coli |
| 1061 | AAAcaaAGTGAGCAAA | Adenomatous polyposis coli |
| 1062 | CAAAGtgaGCAAAGACAA | Adenomatous polyposis coli |
| 1065 | CAAAGacAATCAAGGAA | Adenomatous polyposis coli |
| 1067 | CAAtcaaGGAATCAAAG | Adenomatous polyposis coli |
| 1071 | CAAAGtACAACTTATC | Adenomatous polyposis coli |
| 1079 | ACTGagAGCACTGATG | Adenomatous polyposis coli |
| 1082 | ACTGAtgATAAACACCT | Adenomatous polyposis coli |
| 1084 | GATAaacACCTCAAGTT | Adenomatous polyposis coli |
| 1086 | CACCTcAAGTTCCAAC | Adenomatous polyposis coli |
| 1093 | TTTGgACAGCAGGAA | Adenomatous polyposis coli |
| 1098 | TGTgtTTCTCCATAC | Adenomatous polyposis coli |
| 1105 | CGGgGAGCCAATGG | Thyroid cancer |
| 1110 | TCAGAaACAAATCGAG | Adenomatous polyposis coli |
| 1121 | ATTAAtcaaAATGTAAGCC | Adenomatous polyposis coli |
| 1131 | CAAAGAGATGACTA | Adenomatous polyposis coli |
| 1134 | GACTAtGAAGATGATA | Adenomatous polyposis coli |
| 1137 | GATgataaGCCTACCAAT | Adenomatous polyposis coli |
| 1146 | CGTTAcTCTGAAGAAG | Adenomatous polyposis coli |
| 1154 | GAAAGaagaaGAGAGACCAA | Adenomatous polyposis coli |

| | | |
|------|----------------------|----------------------------|
| 1155 | GAAGaagaGAGACCAACA | Adenomatous polyposis coli |
| 1156 | GAAgagaGACCAACAAA | Adenomatous polyposis coli |
| 1168 | GAAGagaaACGTCATGTG | Adenomatous polyposis coli |
| 1178 | GATTAtagttAAATATGCCA | Adenomatous polyposis coli |
| 1181 | TTAAaATATGCCACA | Adenomatous polyposis coli |
| 1184 | GCCacagaTATTCCCTCA | Adenomatous polyposis coli |
| 1185 | ACAgatATTCCCTTC | Adenomatous polyposis coli |
| 1190 | TCACAgAAACAGTCAT | Adenomatous polyposis coli |
| 1192 | AAAcaGTCATTTCA | Adenomatous polyposis coli |
| 1198 | TCAaaGAGTTCATCT | Adenomatous polyposis coli |
| 1207 | AAAACGAAACATATG | Adenomatous polyposis coli |
| 1208 | ACCgaacATATGTCTTC | Adenomatous polyposis coli |
| 1210 | CATatGTCTCAAGC | Adenomatous polyposis coli |
| 1233 | CCAAGtTCTGCACAGA | Adenomatous polyposis coli |
| 1249 | TGCAaaGTTCTTCTA | Adenomatous polyposis coli |
| 1259 | ATAcAGACTTATTGT | Adenomatous polyposis coli |
| 1260 | CAGACttATTGTGTAGA | Adenomatous polyposis coli |
| 1268 | CCAaTATGTTTTTC | Adenomatous polyposis coli |
| 1275 | AGTtCATTATCATC | Adenomatous polyposis coli |
| 1294 | CAGGAaGCAGATTCTG | Adenomatous polyposis coli |
| 1301 | ACCCtGCAAATAGCA | Adenomatous polyposis coli |
| 1306 | GAAAAtaaaAGAAAAGATT | Adenomatous polyposis coli |

| | | |
|------|--------------------------------|----------------------------|
| 1307 | ATAaAAGAAAAGAT | Adenomatous polyposis coli |
| 1308 | AAAaaaAGATTGGAAC | Adenomatous polyposis coli |
| 1308 | AAAGAaaagaTTGGAAC | Adenomatous polyposis coli |
| 1318 | GATCcTGTGAGCGAA | Adenomatous polyposis coli |
| 1320 | GTGAGcGAAGTTCCAG | Adenomatous polyposis coli |
| 1323 | GTTCcAGCAGTGTCA | Adenomatous polyposis coli |
| 1329 | CACCCttagaaccAAATCCAGCA | Adenomatous polyposis coli |
| 1336 | AGACtgCAGGGTTCTA | Adenomatous polyposis coli |
| 1338 | CAGgGTTCTAGTTT | Adenomatous polyposis coli |
| 1340 | TCTAgTTTATCTTCA | Adenomatous polyposis coli |
| 1342 | TTATCttCAGAATCA | Adenomatous polyposis coli |
| 1352 | GTTgAATTTCCTTC | Adenomatous polyposis coli |
| 1361 | CCCTcCAAAAGTGGT | Adenomatous polyposis coli |
| 1364 | AGTggtgCTCAGACACC | Adenomatous polyposis coli |
| 1371 | AGTCCacCTGAACACTA | Adenomatous polyposis coli |
| 1372 | CCACCCtGAACACTATG | Adenomatous polyposis coli |
| 1376 | TATGttCAGGAGACCC | Adenomatous polyposis coli |
| 1394 | GATAgtTTGAGAGTC | Adenomatous polyposis coli |
| 1401 | ATTGCcAGCTCCGTT | Adenomatous polyposis coli |
| 1415 | AGTGGcATTATAAGCC | Adenomatous polyposis coli |
| 1426 | AGCCcTGGACAAACC | Adenomatous polyposis coli |
| 1427 | CCTGGaCAAACCATGC | Adenomatous polyposis coli |

| | | |
|------|-----------------------|----------------------------|
| 1431 | ATGCcACCAAGCAGA | Adenomatous polyposis coli |
| 1454 | AAAAAAtAAAGCACCTA | Adenomatous polyposis coli |
| 1461 | GAAaAGAGAGAGAG | Adenomatous polyposis coli |
| 1463 | AGAgagaGTGGACCTAA | Adenomatous polyposis coli |
| 1464 | GAGAgTGGACCTAAG | Adenomatous polyposis coli |
| 1464 | GAGAgGGACCTAAGC | Adenomatous polyposis coli |
| 1464 | GAGagTGGACCTAAG | Adenomatous polyposis coli |
| 1492 | GCCaCGGAAAGTAC | Adenomatous polyposis coli |
| 1493 | ACGGAAAGTACTCCAG | Adenomatous polyposis coli |
| 1497 | CCAgATGGATTTC | Adenomatous polyposis coli |
| 1503 | TCAAtccaGCCTGAGTGC | Adenomatous polyposis coli |
| 1522 | TTAagaataaTGCCTCCAGT | Adenomatous polyposis coli |
| 1536 | GAAACagAATCAGAGCA | Adenomatous polyposis coli |
| 1545 | TCAAAAtgaaaACCAAGAGAA | Adenomatous polyposis coli |
| 1547 | GAAaACCAAGAGAA | Adenomatous polyposis coli |
| 1550 | GAGAaagaGGCAGAAAAA | Adenomatous polyposis coli |
| 1577 | GAATgtATTATTCTG | Adenomatous polyposis coli |
| 1594 | CCAGCcCAGACTGCTT | Adenomatous polyposis coli |
| 1596 | CAGACTGCTCAAAAT | Adenomatous polyposis coli |
| 1823 | TTCAaTGATAAGCTC | Adenomatous polyposis coli |
| 1859 | AATGAttctTTGAGTTCTC | Adenomatous polyposis coli |
| 1941 | CCAGAcagaGGGCAGCAA | Desmoid tumours |
| 1957 | GAAaATACTCCAGT | Adenomatous |

| | | |
|------|-------------------------------|----------------------------|
| | | polyposis coli |
| 1980 | AACaATAAAGAAAA | Adenomatous polyposis coli |
| 1985 | GAACCTATCAAAGAGA | Adenomatous polyposis coli |
| 1986 | CCTaTCAAAGAGAC | Adenomatous polyposis coli |
| 1998 | GAACcAAGTAAACCT | Adenomatous polyposis coli |
| 2044 | AGCTCcGCAATGCCAA | Adenomatous polyposis coli |
| 2556 | TCATCcttcctcGAGTAAGCAC | Adenomatous polyposis coli |
| 2643 | CTAATttatCAAATGGCAC | Adenomatous polyposis coli |

TABLE VI:
SMALL INSERTIONS

| Codon | Insertion | Phenotype |
|-------|-----------|----------------------------|
| 157 | T | Adenomatous polyposis coli |
| 170 | AGAT | Adenomatous polyposis coli |
| 172 | T | Adenomatous polyposis coli |
| 199 | G | Adenomatous polyposis coli |
| 243 | AG | Adenomatous polyposis coli |
| 266 | T | Adenomatous polyposis coli |
| 357 | A | Adenomatous polyposis coli |
| 405 | C | Adenomatous polyposis coli |
| 413 | T | Adenomatous polyposis coli |
| 416 | A | Adenomatous polyposis coli |
| 457 | G | Adenomatous polyposis coli |
| 473 | A | Adenomatous polyposis coli |
| 503 | ATTC | Adenomatous polyposis coli |
| 519 | C | Adenomatous polyposis coli |
| 528 | A | Adenomatous polyposis coli |
| 561 | A | Adenomatous polyposis coli |
| 608 | A | Adenomatous polyposis coli |
| 620 | CT | Adenomatous polyposis coli |
| 621 | A | Adenomatous polyposis coli |
| 623 | TTAC | Adenomatous polyposis coli |
| 627 | A | Adenomatous polyposis coli |
| 629 | A | Adenomatous polyposis coli |
| 636 | GT | Adenomatous polyposis coli |
| 639 | A | Adenomatous polyposis coli |

| | | |
|------|------|----------------------------|
| 704 | T | Adenomatous polyposis coli |
| 740 | ATGC | Adenomatous polyposis coli |
| 764 | T | Adenomatous polyposis coli |
| 779 | TT | Adenomatous polyposis coli |
| 807 | AT | Adenomatous polyposis coli |
| 827 | AT | Adenomatous polyposis coli |
| 831 | A | Adenomatous polyposis coli |
| 841 | CTTA | Adenomatous polyposis coli |
| 865 | CT | Adenomatous polyposis coli |
| 865 | AT | Adenomatous polyposis coli |
| 900 | TG | Adenomatous polyposis coli |
| 921 | G | Adenomatous polyposis coli |
| 927 | A | Adenomatous polyposis coli |
| 935 | A | Adenomatous polyposis coli |
| 936 | C | Adenomatous polyposis coli |
| 975 | A | Adenomatous polyposis coli |
| 985 | T | Adenomatous polyposis coli |
| 997 | A | Adenomatous polyposis coli |
| 1010 | TA | Adenomatous polyposis coli |
| 1085 | C | Adenomatous polyposis coli |
| 1085 | AT | Adenomatous polyposis coli |
| 1095 | A | Adenomatous polyposis coli |
| 1100 | TTT | Adenomatous polyposis coli |
| 1107 | GGAG | Adenomatous polyposis coli |
| 1120 | G | Adenomatous polyposis coli |
| 1166 | A | Adenomatous polyposis coli |
| 1179 | T | Adenomatous polyposis coli |
| 1187 | A | Adenomatous polyposis coli |
| 1211 | T | Adenomatous polyposis coli |
| 1256 | A | Adenomatous polyposis coli |
| 1265 | T | Adenomatous polyposis coli |
| 1267 | GATA | Adenomatous polyposis coli |
| 1268 | T | Adenomatous polyposis coli |
| 1301 | A | Adenomatous polyposis coli |
| 1301 | C | Adenomatous polyposis coli |
| 1323 | A | Adenomatous polyposis coli |
| 1342 | T | Adenomatous polyposis coli |
| 1382 | T | Adenomatous polyposis coli |
| 1458 | GTAG | Adenomatous polyposis coli |
| 1463 | AG | Adenomatous polyposis coli |
| 1488 | T | Adenomatous polyposis coli |
| 1531 | A | Adenomatous polyposis coli |
| 1533 | T | Adenomatous polyposis coli |
| 1554 | A | Adenomatous polyposis coli |

| | | |
|------|-------|----------------------------|
| 1555 | A | Adenomatous polyposis coli |
| 1556 | T | Adenomatous polyposis coli |
| 1563 | GACCT | Adenomatous polyposis coli |
| 1924 | AA | Desmoid tumours |

TABLE VII: SMALL INSERTIONS/DELETIONS

| Location/ codon | Deletion | Insertion | Phenotype |
|--------------------|--------------------|-------------|----------------------------|
| 538 | GAAGAcTTACAGCAGG | gaa | Adenomatous polyposis coli |
| 620 | CTTACttaCCGGAGCCAG | ct | Adenomatous polyposis coli |
| 728 | AATctcatGGCAAATAGG | Ttgcagcttaa | Adenomatous polyposis coli |
| 971 | GATGgtTATGGTAAAA | taa | Adenomatous polyposis coli |

TABLE VIII:
GROSS DELETIONS

| | |
|-------------------------------------|----------------------------|
| 2 kb including ex. 11 | Adenomatous polyposis coli |
| 3 kb I10E11-1.5 kb to I12E13-170 bp | Adenomatous polyposis coli |
| 335 bp nt. 1409-1743 ex. 11-13 | Adenomatous polyposis coli |
| 6 kb incl. ex. 14 | Adenomatous polyposis coli |
| 817 bp I13E14-679 to I13E14+138 | Adenomatous polyposis coli |
| ex. 11-15M | Adenomatous polyposis coli |
| ex. 11-3'UTR | Adenomatous polyposis coli |
| ex. 15A - ex. 15F | Adenomatous polyposis coli |
| ex. 4 | Adenomatous polyposis coli |
| ex. 7, 8 and 9 | Adenomatous polyposis coli |
| ex. 8 to beyond ex. 15F | Adenomatous polyposis coli |
| ex. 8 - ex. 15F | Adenomatous polyposis coli |
| ex. 9 | Adenomatous polyposis coli |
| >10mb (del 5q22) | Adenomatous polyposis coli |

TABLE IX:
GROSS INSERTIONS AND DUPLICATIONS

| Description | Phenotype |
|-----------------------------|----------------------------|
| Insertion of 14 bp nt. 3816 | Adenomatous polyposis coli |

| | |
|--|----------------------------|
| Insertion of 22 bp nt. 4022 | Adenomatous polyposis coli |
| Duplication of 43 bp cd. 1295 | Adenomatous polyposis coli |
| Insertion of 337 bp of Alu I sequence cd. 1526 | Desmoid tumours |

TABLE X:
COMPLEX REARRANGEMENTS (INCLUDING INVERSIONS)

| | |
|--|----------------------------|
| A-T nt. 4893 Q1625H, Del C nt. 4897 cd. 1627 | Adenomatous polyposis coli |
| Del 1099 bp I13E14-728 to E14I14+156, ins 126 bp | Adenomatous polyposis coli |
| Del 1601 bp E14I14+27 to E14I14+1627, ins 180 bp | Adenomatous polyposis coli |
| Del 310 bp, ins. 15 bp nt. 4394, cd 1464 | Adenomatous polyposis coli |
| Del A and T cd. 1395 | Adenomatous polyposis coli |
| Del TC nt. 4145, Del TGT nt. 4148 | Adenomatous polyposis coli |
| Del. T, nt. 983, Del. 70 bp, nt. 985 | Adenomatous polyposis coli |
| Del. nt. 3892-3903, ins ATTT | Adenomatous polyposis coli |

TABLE XI:
DIAGNOSTIC APPLICATIONS

| Cancer Type | Marker | Application | Reference |
|----------------|---|---|--|
| Breast | Her2/Neu Detection – polymorphism at codon 655 (GTC/valine to ATC/isoleucine [Val(655)Ile]) | Using methods described herein, design second primer such that after PCR, and digestion with restriction enzyme, a 5' overhang containing DNA sequence for codon 655 of Her2/Neu is generated. Her2/Neu can be detected and quantified as a possible marker for breast cancer. Methods described herein can detect both mutant allele and normal allele, even when mutant allele is small fraction of total DNA. Herceptin therapy for breast cancer is based upon screening for Her2. The earlier the mutant allele can be detected, the faster therapy can be provided. | D. Xie <i>et al.</i> , <i>J. Natl. Cancer Institute</i> , 92, 412 (2000) K.S. Wilson <i>et al.</i> , <i>Am. J. Pathol.</i> , 161, 1 171 (2002) L. Newman, <i>Cancer Control</i> , 9, 473 (2002) [1] |
| Breast/Ovarian | Hypermethylati on of BRCA1 | Methods described herein can be used to differentiate between tumors resulting from inherited BRCA1 mutations and those from non- inherited abnormal methylation of the gene | M. Esteller <i>et al.</i> , <i>New England Jnl Med.</i> , 344, 539 (2001) |

| Cancer Type | Marker | Application | Reference |
|---------------|--|--|--|
| Bladder | Microsatellite analysis of free tumor DNA in Urine, Serum and Plasma | Methods described herein can be applied to microsatellite analysis and FGFR3 mutation analysis for detection of bladder cancer. Methods described herein provide a non-invasive method for detection of bladder cancer. | W.G. Bas <i>et al.</i> , <i>Clinical Cancer Res.</i> , 9,257 (2003) M. Utting <i>et al.</i> , <i>Clinical Cancer Res.</i> , 8,35 (2002) L. Mao, D.Sidransky <i>et al.</i> , <i>Science</i> , 271, 669 (1996) |
| Lung | Microsatellite analysis of DNA from sputum | Methods described herein can be used to detect mutations in sputum samples, and can markedly boost the accuracy of preclinical lung cancer screening | T.Liloglou <i>et al.</i> , <i>Cancer Research</i> , 61, 1624, (2001) M. Tockman <i>et al.</i> , <i>Cancer Control</i> , 7, 19 (2000) Field <i>et al.</i> , <i>Cancer Research</i> , 59, 2690 (1999) |
| Cervical | Analysis of HPV genotype | Methods described herein can be used to detect HPV genotype from a cervical smear preparation. | N. Munoz <i>et al.</i> , <i>New England Jnl Med.</i> , 348, 518 (2003) |
| Head and Neck | Tumor specific alterations in exfoliated oral mucosal cells (microsatellite markers) | Methods described herein can be used to detect any of 23 microsatellite markers, which are associated with Head and Neck Squamous Cell Carcinoma (HNSCC). | M. Spafford <i>et al.</i> <i>Clinical Cancer Research</i> , 17, 607 (2001) A. El-Naggar <i>et al.</i> , <i>J. Mol. Diag.</i> , 3,164 (2001) |

| Cancer Type | Marker | Application | Reference |
|-------------|---|--|---|
| Colorectal | Screening for mutation in K-ras2 and APC genes. | Methods described herein can be used to detect K-ras 2 mutations, which can be used as a prognostic indicator for colorectal cancer. APC (see Example 5). | B. Ryan <i>et al.</i> <i>Gut</i> , 52, 101 (2003) |
| Prostate | GSTP1 Hypermethylation | Methods described herein can be used to detect GSTP1 hypermethylation in urine from patients with prostate cancer; this can be a more accurate indicator than PSA. | P. Cairns <i>et al.</i> <i>Clin. Can. Res.</i> , 7, 2727 (2001) |

HIV

| | | | |
|---------------------------|--|--|--|
| Antiretroviral resistance | Screening individuals for mutations in HIV virus – e.g. 154V mutation or CCR5 Δ 32 allele. | Methods described herein can be used for detection of mutations in the HIV virus. Treatment outcomes are improved in individuals receiving antiretroviral therapy based upon resistance screening. | J. Durant <i>et al.</i> <i>The Lancet</i> , 353, 2195 (1999) |
|---------------------------|--|--|--|

CARDIOLOGY

| | | | |
|--------------------------|---|---|---|
| Congestive Heart Failure | Synergistic polymorphisms of betal and alpha2c adrenergic receptors | Methods described herein can be used to genotype these loci and may help identify people who are at a higher risk of heart failure. | K. Small <i>et al.</i> <i>New Eng. Jnl. Med.</i> , 347, 1135 (2002) |
|--------------------------|---|---|---|

[0450] Having now fully described the invention, it will be understood by those of skill in the art that the invention can be performed with a wide and equivalent range of conditions, parameters, and the like, without affecting the spirit or scope of the invention or any embodiment thereof.

[0451] All documents, e.g., scientific publications, patents and patent publications recited herein are hereby incorporated by reference in their entirety to the same extent as if each individual document was specifically and individually

indicated to be incorporated by reference in its entirety. Where the document cited only provides the first page of the document, the entire document is intended, including the remaining pages of the document.

WHAT IS CLAIMED IS:

1. A method for determining a sequence of a locus of interest, said method comprising:

- (a) replicating a region of DNA comprising a locus of interest from a template polynucleotide by using a first and a second primer, wherein the second primer contains a sequence that generates a recognition site for a restriction enzyme such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest;
- (b) digesting the DNA with the restriction enzyme that recognizes the recognition site generated by the second primer to create a DNA fragment;
- (c) incorporating a nucleotide into the digested DNA of (b) by using the 5' overhang containing the locus of interest as a template; and
- (d) determining the sequence of the locus of interest by determining the sequence of the DNA of (c).

2. A method for determining a sequence of a locus of interest; said method comprising:

- (a) amplifying a locus of interest on a template DNA using a first and second primers, wherein the second primer contains a recognition site for a restriction enzyme such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest;
- (b) digesting the amplified DNA with the restriction enzyme that recognizes the recognition site on the second primer;
- (c) incorporating a nucleotide into the digested DNA of (b) by using the 5' overhang containing the locus of interest as a template; and
- (d) determining the sequence of the locus of interest by determining the sequence of the DNA of (c).

3. The method of claim 2, wherein the template DNA is obtained from a source selected from the group consisting of a bacterium, fungus, virus, protozoan, plant, animal and human.

4. The method of claim 2, wherein the template DNA is obtained from a human source.

5. The method of claim 2, wherein the template DNA is obtained from a sample selected from the group consisting of a cell, tissue, blood, serum, plasma, urine, spinal fluid, lymphatic fluid, semen, vaginal secretion, ascitic fluid, saliva, mucosa secretion, peritoneal fluid, fecal matter, or body exudates.

6. The method of claim 2, wherein the amplification in (a) comprises polymerase chain reaction (PCR).

7. The method of claim 2, wherein the restriction enzyme cuts DNA at the recognition site.

8. The method of claim 7, wherein a 5' region of the second primer does not anneal to the template DNA.

9. The method of claim 7, wherein a 5' region of the first primer does not anneal to the template DNA.

10. The method of claim 7, wherein the restriction enzyme is selected from the group consisting of BsaJ I, Bssk I, Dde I, EcoN I, Fnu4H I, and Hinf I.

11. The method of claim 2, wherein the restriction enzyme cuts DNA at a distance from the recognition site.

12. The method of claim 11, wherein a 5' region of the second primer does not anneal to the template DNA.

13. The method of claim 11, wherein a 5' region of the first primer does not anneal to the template DNA.

14. The method of claim 12, wherein an annealing length of the 3' region of the second primer is selected from the group consisting of 25-20, 20-15, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, and less than 4 bases.

15. The method of claim 2, wherein an annealing temperature for cycle 1 of PCR is about the melting temperature of the portion of the 3' region of the second primer that anneals to the template DNA.

16. The method of claim 15, wherein an annealing temperature for cycle 2 of PCR is about the melting temperature of the portion of the 3' region of the first primer that anneals to the template DNA.

17. The method of claim 16, wherein an annealing temperature for the remaining cycles of PCR is at about the melting temperature of the entire second primer.

18. The method of claim 2, wherein the 3' end of the second primer is adjacent to the locus of interest.

19. The method of claim 11, wherein the recognition site is for a Type IIS restriction enzyme.

20. The method of claim 19, wherein the Type IIS restriction enzyme is selected from the group consisting of: Alw I, Alw26 I, Bbs I, Bbv I, BceA I,

Bmr I, Bsa I, Bst71 I, BsmA I, BsmB I, BsmF I, BspM I, Ear I, Fau I, Fok I, Hga I, Ple I, Sap I, SSfaN I, and Sthi32 I.

21. The method of claim 19, wherein the Type IIS restriction enzyme is BceA I or BsmF I.

22. The method of claim 2, wherein the first primer contains a recognition site for a restriction enzyme that is different from the recognition site for the restriction enzyme on the second primer.

23. The method of claim 22, further comprising digesting the DNA of (c) with a restriction enzyme that recognizes the recognition site on the first primer.

24. The method of claim 2, wherein the first or second primer contains a tag at the 5' terminus.

25. The method of claim 2, wherein the first primer contains a tag at the 5' terminus.

26. The method of claim 27, wherein the tag is used to separate the amplified DNA from the template DNA.

27. The method of claim 26, wherein the tag is used to separate the amplified DNA containing the incorporated nucleotide from the amplified DNA that does not contain the incorporated nucleotide.

28. The method of claim 24, wherein the tag is selected from the group consisting of: radioisotope, fluorescent reporter molecule, chemiluminescent reporter molecule, antibody, antibody fragment, hapten, biotin, derivative of

biotin, photobiotin, iminobiotin, digoxigenin, avidin, enzyme, acridinium, sugar, enzyme, apoenzyme, homopolymeric oligonucleotide, hormone, ferromagnetic moiety, paramagnetic moiety, diamagnetic moiety, phosphorescent moiety, luminescent moiety, electrochemiluminescent moiety, chromatic moiety, moiety having a detectable electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity, and combinations thereof.

29. The method of claim 24, wherein the tag is biotin.

30. The method of claim 29, wherein the biotin tag is used to separate amplified DNA from the template DNA using a streptavidin matrix.

31. The method of claim 30, wherein the streptavidin matrix is coated on wells of a microtiter plate.

32. The method of claim 2, wherein the incorporation of a nucleotide in (c) is by a DNA polymerase selected from the group consisting of E. coli DNA polymerase, Klenow fragment of E. coli DNA polymerase I, T7 DNA polymerase, T4 DNA polymerase, Taq polymerase, Pfu DNA polymerase, Vent DNA polymerase and sequenase.

33. The method of claim 2, wherein the incorporation of a nucleotide in (c) comprises incorporation of a labeled nucleotide.

34. The method of claim 2, wherein the incorporation of a nucleotide in (c) further comprises incorporation of an unlabeled nucleotide.

35. The method of claim 33, wherein the labeled nucleotide is selected from the group consisting of a dideoxynucleotide and deoxynucleotide.

36. The method of claim 33, wherein the labeled nucleotide is labeled with a molecule selected from the group consisting of radioactive molecule, fluorescent molecule, antibody, antibody fragment, hapten, carbohydrate, biotin, derivative of biotin, phosphorescent moiety, luminescent moiety, electrochemiluminescent moiety, chromatic moiety, and moiety having a detectable electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity.

37. The method of claim 33, wherein the labeled nucleotide is labeled with a fluorescent molecule.

38. The method of claim 37, further comprising incorporation of an unlabeled nucleotide.

39. The method of claim 2, wherein the determination of the sequence of the locus of interest in (d) comprises detecting the nucleotide.

40. The method of claim 33, wherein the determination of the sequence of the locus of interest in (d) comprises detecting labeled nucleotide.

41. The method of claim 40, wherein the detection is by a method selected from the group consisting of gel electrophoresis, polyacrylamide gel electrophoresis, fluorescence detection, sequencing, ELISA, mass spectrometry, fluorometry, hybridization, microarray, and Southern Blot.

42. The method of claim 40, wherein the detection method is DNA sequencing.

43. The method of claim 40, wherein the detection method is fluorescence detection.

44. The method of claim 2, wherein the locus of interest is suspected of containing a single nucleotide polymorphism or mutation.

45. The method of claim 2, wherein the method is used for determining sequences of multiple loci of interest concurrently.

46. The method of claim 45, wherein the template DNA comprises multiple loci from a single chromosome.

47. The method of claim 45, wherein the template DNA comprises multiple loci from different chromosomes.

48. The method of claim 45, wherein the loci of interest on template DNA are amplified in one reaction.

49. The method of claim 45, wherein each of the loci of interest on template DNA is amplified in a separate reaction.

50. The method of claim 49, wherein the amplified DNA are pooled together prior to digestion of the amplified DNA.

51. The method of claim 45, wherein each of the labeled DNA in (c) containing a locus of interest is separated prior to (d).

52. The method of claim 45, wherein at least one of the loci of interest is suspected of containing a single nucleotide polymorphism or a mutation.

53. A method for determining a sequence of a locus of interest, said method comprising:

- (a) amplifying a locus of interest on a template DNA using a first and second primers, wherein the second primer contains a portion of a recognition site for a restriction enzyme, wherein a full recognition site for the restriction enzyme is generated upon amplification of the template DNA such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest;
- (b) digesting the amplified DNA with the restriction enzyme that recognizes the full recognition site generated by the second primer and the template DNA;
- (c) incorporating a nucleotide into the digested DNA of (b) by using the 5' overhang containing the locus of interest as a template; and determining the sequence of the locus of interest by determining the sequence of the DNA of (c).

54. The method of claim 52, wherein the 3' region of the second primer contains a mismatch with the template DNA.

55. The method of claim 53, wherein the mismatch occurs at the last 1, 2 or 3 bases at the 3' region.

56. A method for determining a sequence of a locus of interest, said method comprising:

- (a) amplifying a locus of interest on a template DNA using a first and second primers, wherein the second primer contains a recognition site for a restriction enzyme that cuts DNA at a distance from the recognition site and digestion with the restriction enzyme generates a 5' overhang containing the locus of interest, and wherein the first primer contains a recognition site for a restriction enzyme that is different from the recognition site for the restriction enzyme on the second primer and contains a tag at the 5' end;
- (b) digesting the amplified DNA with the restriction enzyme that recognizes the recognition site on the second primer;

- (c) incorporating a labeled nucleotide into the digested DNA of (b) by using the 5' overhang containing the locus of interest as a template;
- (d) digesting the DNA of (c) with the restriction enzyme that recognizes the recognition site on the first primer; and
- (e) determining the sequence of the locus of interest by determining the sequence of the digested DNA of (d) containing the labeled nucleotide.

57. A method for determining a sequence of a locus of interest, said method comprising:

- (a) amplifying a locus of interest on a template DNA using a first and second primers, wherein the second primer contains a recognition site for a restriction enzyme that cuts DNA at a distance from the recognition site and digestion with the restriction enzyme generates a 5' overhang containing the locus of interest, wherein the first primer contains a recognition site for a restriction enzyme that is different from the recognition site for the restriction enzyme on the second primer, and contains a tag at the 5' end, and wherein the annealing temperature for cycle 1 of PCR is at about the melting temperature of the portion of the 3' region of the second primer that anneals to the template DNA, the annealing temperature for cycle 2 of PCR is at about the melting temperature of the portions of the 3' region of the first primer that anneals to the template DNA, and the annealing temperature for the remaining cycles is at about the melting temperature of the entire second primer;
- (b) digesting the amplified DNA with the restriction enzyme that recognizes the recognition site on the second primer;
- (c) incorporating a labeled nucleotide into the digested DNA of (b) by using the 5' overhang containing the locus of interest as a template;
- (d) digesting the DNA of (c) with the restriction enzyme that recognizes the recognition site on the first primer; and
- (e) determining the sequence of the locus of interest by determining the sequence of the digested DNA of (d) containing the labeled nucleotide.

58. The method of claim 56, wherein the tag is used to separate the amplified DNA from the template DNA.

59. A kit for use in any of the methods of claims 1 to 58 comprising a set of primers used in the method, wherein the second primer contains a sequence that generates a recognition site for a restriction enzyme such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest, and a set of instructions.

60. A DNA fragment containing a locus of interest to be sequenced and a recognition site for a restriction enzyme, wherein digestion with the restriction enzyme creates a 5' overhang on the DNA fragment, and wherein the locus of interest and the restriction enzyme recognition site are in relationship to each other such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest.

61. A method for determining a sequence of alleles of a locus of interest, said method comprising:

- (a) amplifying alleles of a locus of interest on a template DNA using a first and second primers, wherein the second primer contains a recognition site for a restriction enzyme such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest;
- (b) digesting the amplified DNA with the restriction enzyme that recognizes the recognition site on the second primer;
- (c) incorporating nucleotides into the digested DNA of (b), wherein
 - (i) a nucleotide that terminates elongation, and is complementary to the locus of interest of an allele, is incorporated into the 5' overhang of said allele, and

- (ii) a nucleotide complementary to the locus of interest of a different allele is incorporated into the 5' overhang of said different allele, and said terminating nucleotide, which is complementary to a nucleotide in the 5' overhang of said different allele, is incorporated into the 5' overhang of said different allele;
- (d) determining the sequence of the alleles of a locus of interest by determining the sequence of the DNA of (c).

62. A method for determining a sequence of alleles of a locus of interest, said method comprising:

- (a) amplifying a locus of interest on a template DNA using a first and second primers, wherein the second primer contains a portion of a recognition site for a restriction enzyme, wherein a full recognition site for the restriction enzyme is generated upon amplification of the template DNA such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest;
- (b) digesting the amplified DNA with the restriction enzyme that recognizes the full recognition site generated by the second primer and the template DNA;
- (c) incorporating nucleotides into the digested DNA of (b), wherein
 - (i) a nucleotide that terminates elongation, and is complementary to the locus of interest of an allele, is incorporated into the 5' overhang of said allele, and
 - (ii) a nucleotide complementary to the locus of interest of a different allele is incorporated into the 5' overhang of said different allele, and said terminating nucleotide, which is complementary to a nucleotide in the 5' overhang of said different allele, is incorporated into the 5' overhang of said different allele;determining the sequence of the locus of interest by determining the sequence of the DNA of (c).

63. A method for determining a sequence of alleles of a locus of interest, said method comprising:

- (a) amplifying a locus of interest on a template DNA using a first and second primers, wherein the second primer contains a recognition site for a restriction enzyme that cuts DNA at a distance from the recognition site and digestion with the restriction enzyme generates a 5' overhang containing the locus of interest, and wherein the first primer contains a recognition site for a restriction enzyme that is different from the recognition site for the restriction enzyme on the second primer and contains a tag at the 5' end;
- (b) digesting the amplified DNA with the restriction enzyme that recognizes the recognition site on the second primer;
- (c) incorporating nucleotides into the digested DNA of (b), wherein
 - (i) a labeled nucleotide that terminates elongation, and is complementary to the locus of interest of an allele, is incorporated into the 5' overhang of said allele, and
 - (ii) a nucleotide complementary to the locus of interest of a different allele is incorporated into the 5' overhang of said different allele, and said terminating nucleotide, which is complementary to a nucleotide in the 5' overhang of said different allele, is incorporated into the 5' overhang of said different allele;
- (d) digesting the DNA of (c) with the restriction enzyme that recognizes the recognition site on the first primer; and
- (e) determining the sequence of the alleles of a locus of interest by determining the sequence of the digested DNA of (d) containing the labeled nucleotide.

64. A method for determining a sequence of alleles of a locus of interest, said method comprising:

- (a) amplifying a locus of interest on a template DNA using a first and second primers, wherein the second primer contains a recognition site for a restriction

enzyme that cuts DNA at a distance from the recognition site and digestion with the restriction enzyme generates a 5' overhang containing the locus of interest, wherein the first primer contains a recognition site for a restriction enzyme that is different from the recognition site for the restriction enzyme on the second primer, and contains a tag at the 5' end, and wherein the annealing temperature for cycle 1 of PCR is at about the melting temperature of the portion of the 3' region of the second primer that anneals to the template DNA, the annealing temperature for cycle 2 of PCR is at about the melting temperature of the portion of the 3' region of the first primer that anneals to the template DNA, and the annealing temperature for the remaining cycles of PCR is at about the melting temperature of the entire second primer;

- (b) digesting the amplified DNA with the restriction enzyme that recognizes the recognition site on the second primer;
- (c) incorporating nucleotides into the digested DNA of (b), wherein
 - (i) a labeled nucleotide that terminates elongation, and is complementary to the locus of interest of an allele, is incorporated into the 5' overhang of said allele, and
 - (ii) a nucleotide complementary to the locus of interest of a different allele is incorporated into the 5' overhang of said different allele, and said terminating nucleotide, which is complementary to a nucleotide in the 5' overhang of said different allele, is incorporated into the 5' overhang of said different allele;
- (d) digesting the DNA of (c) with the restriction enzyme that recognizes the recognition site on the first primer; and
- (e) determining the sequence of the alleles of a locus of interest by determining the sequence of the digested DNA of (d) containing the labeled nucleotide.

65. The method of claim 61, wherein said method is used for determining the sequence of a mutant allele in the presence of normal allele.

66. The method of claim 65, wherein the mutant allele and normal allele are present in a ratio selected from the group consisting of 1:1, 1:2, 1:3, 1:4, 1:5, 1:6-1:10, 1:11-1:20, 1:21-1:30, 1:31-1:40, 1:41-1:50, 1:51-1:60, 1:61-1:70, 1:71-1:80, 1:81-1:90, 1:91:1:100, 1:101-1:200, 1:250, 1:251-1:300, 1:301-1:400, 1:401-1:500, 1:501-1:600, 1:601-1:700, 1:701-1:800, 1:801-1:900, 1:901-1:1000, 1:1001-1:2000, 1:2001-1:3000, 1:3001-1:4000, 1:4001-1:5000, 1:5001-1:6000, 1:6001-1:7000, 1:7001-1:8000, 1:8001-1:9000, 1:9001-1:10,000; 1:10,001-1:20,000, 1:20,001:1:30,000, 1:30,001-1:40,000, 1:40,001-1:50,000, and greater than 1:50,000.

67. The method of claim 65, wherein said mutant allele is of a gene selected from the group consisting of BRCA1, BRCA2, MSH6, MSH2, MLH1, RET, PTEN, ATM, H-RAS, p53, ELAC2, CDH1, APC, AR, PMS2, MLH3, CYP1A1, GSTP1, GSTM1, AXIN2, CYP19, MET, NAT1, CDKN2A, NQ01, trc8, RAD51, PMS1, TGFBR2, VHL, MC4R, POMC, NROB2, UCP2, PCSK1, PPARG, ADRB2, UCP3, glur1, cart, SORBS1, LEP, LEPR, SIM1, TNF, IL-6, IL-1, IL-2, IL-3, IL1A, TAP2, THPO, THRIB, NBS1, RBM15, LIF, MPL, RUNX1, Her-2, glucocorticoid receptor, estrogen receptor, thyroid receptor, p21, p27, K-RAS, N-RAS, retinoblastoma protein, Wiskott-Aldrich (WAS) gene, Factor V Leiden, Factor II (prothrombin), methylene tetrahydrofolate reductase, cystic fibrosis, LDL receptor, HDL receptor, superoxide dismutase gene, and SHOX gene.

FIG. 1A

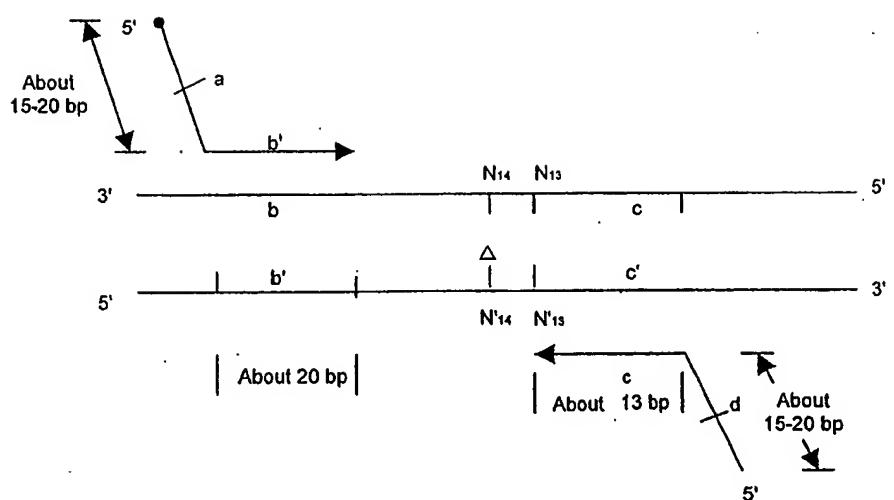


FIG. 1B

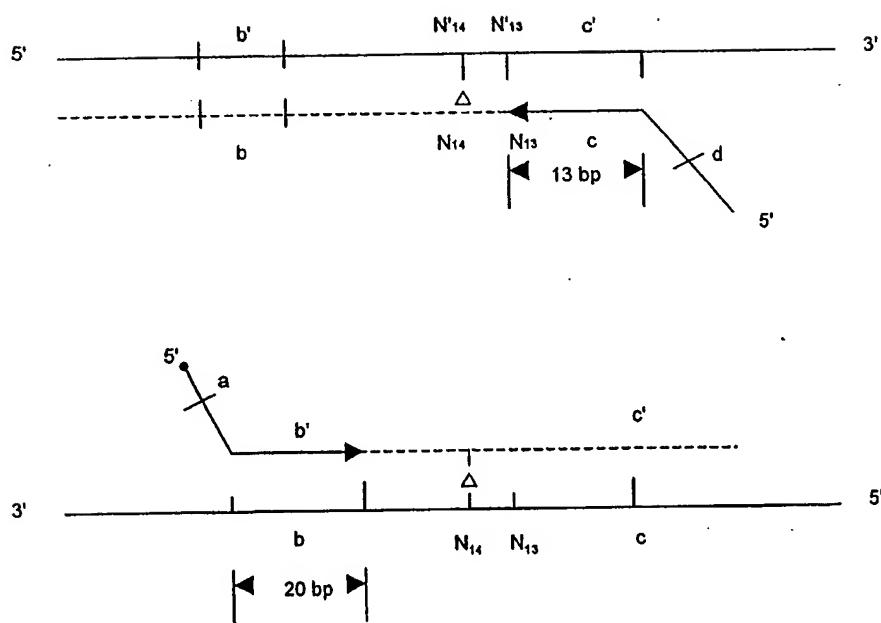


FIG. 1C

First Primer

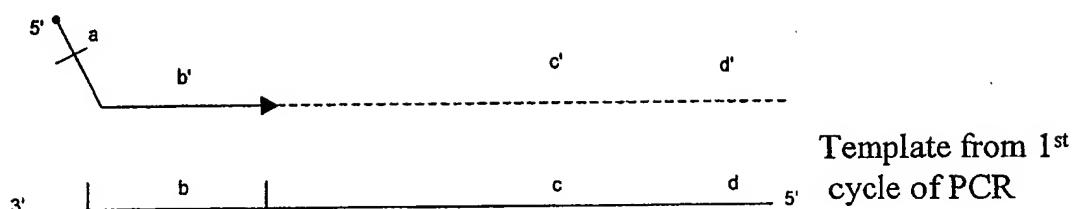
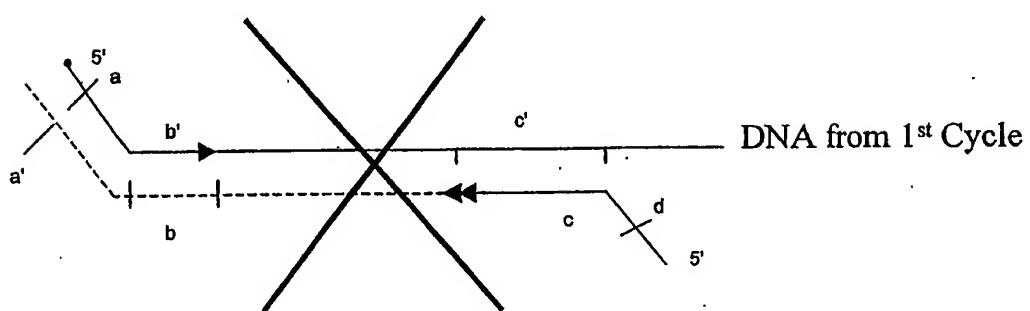
Template from 1st
cycle of PCRDNA from 1st Cycle

FIG. 1D

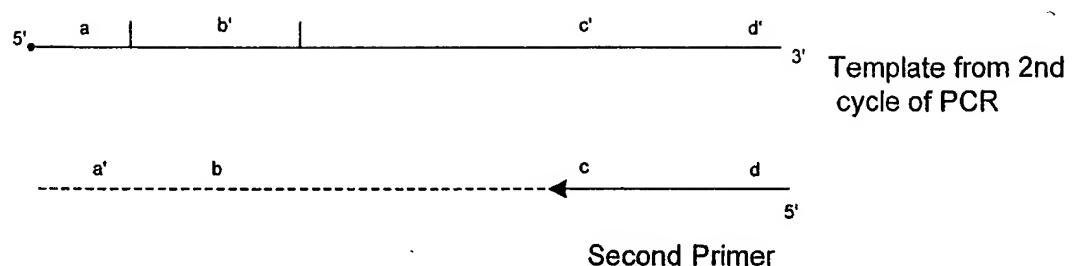


FIG. 1E

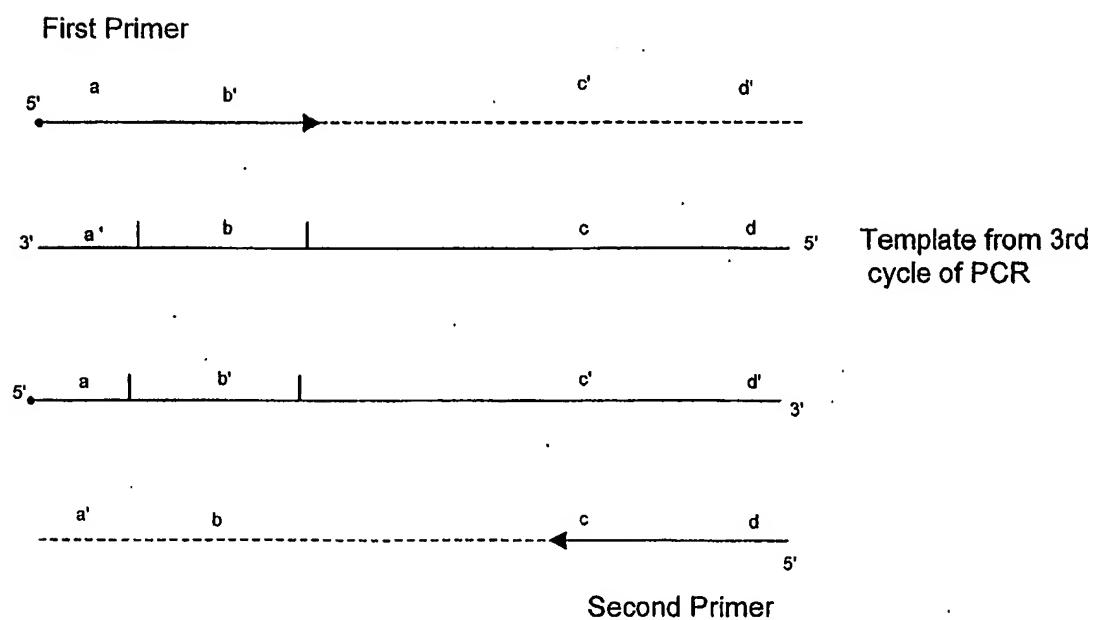


FIG. 1F

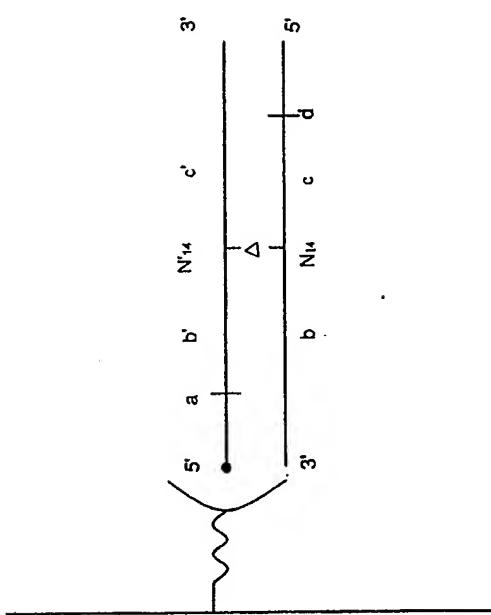


FIG. 1G

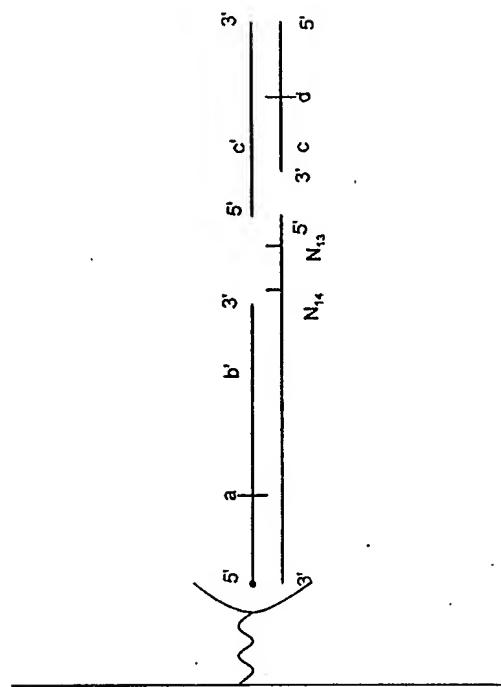


FIG. 1H

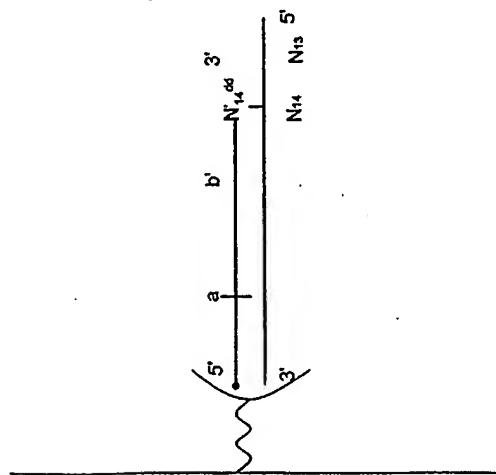


FIG. 1I

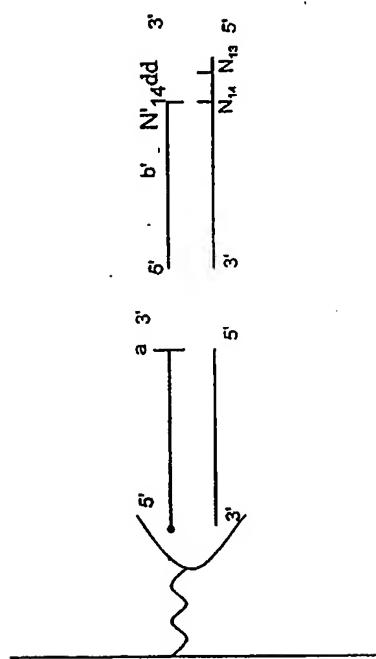


FIG. 2

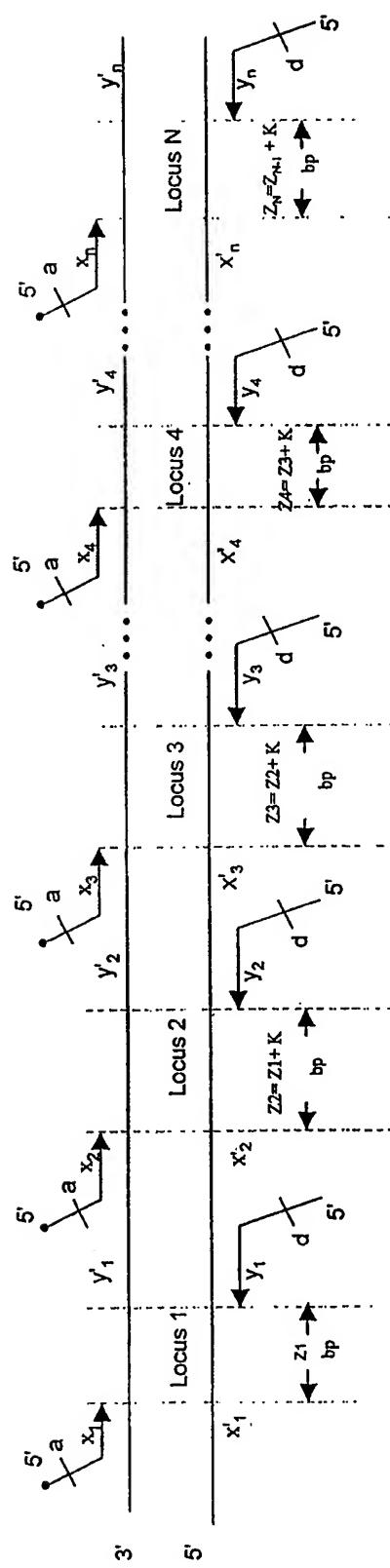


FIG. 3A

FIG. 3B

FIG. 3C

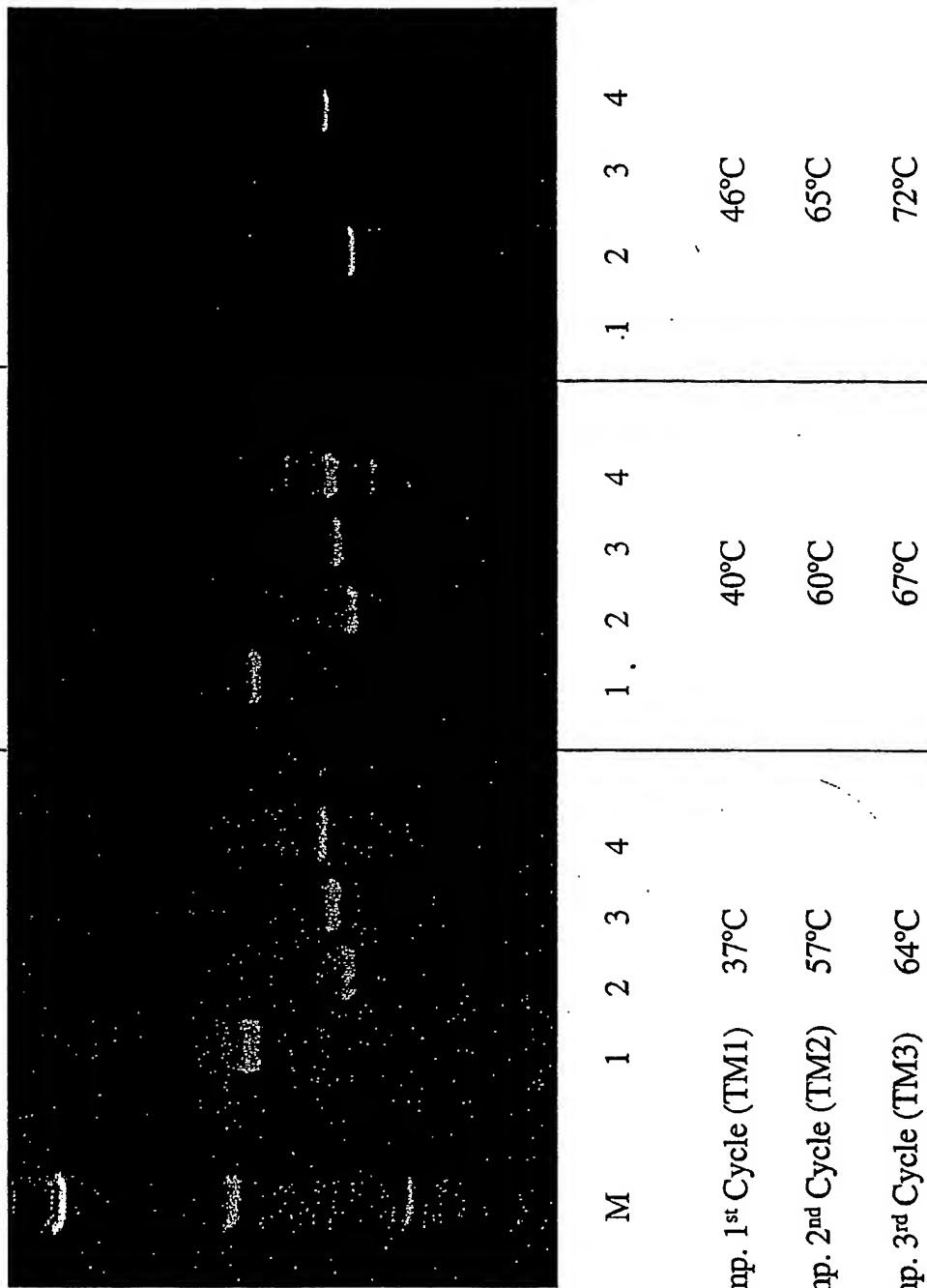


FIG. 4A

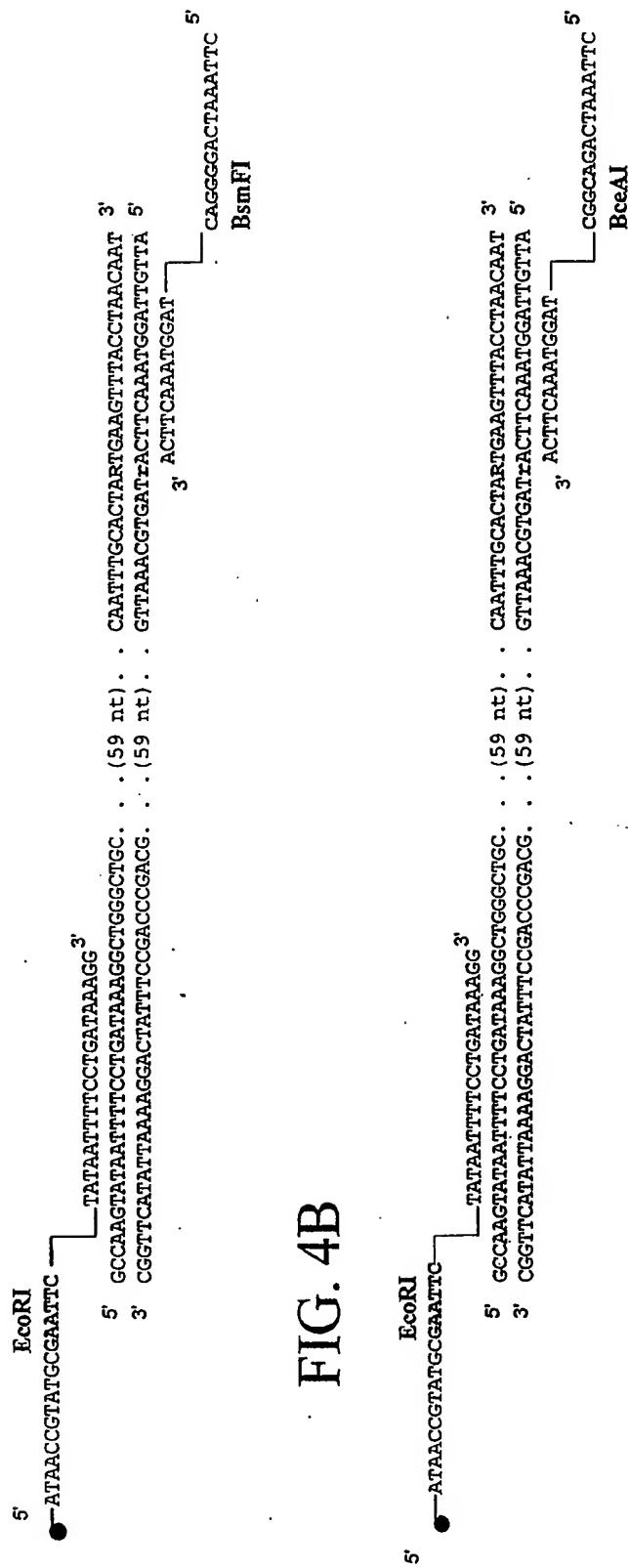


FIG. 4B



FIG. 4C

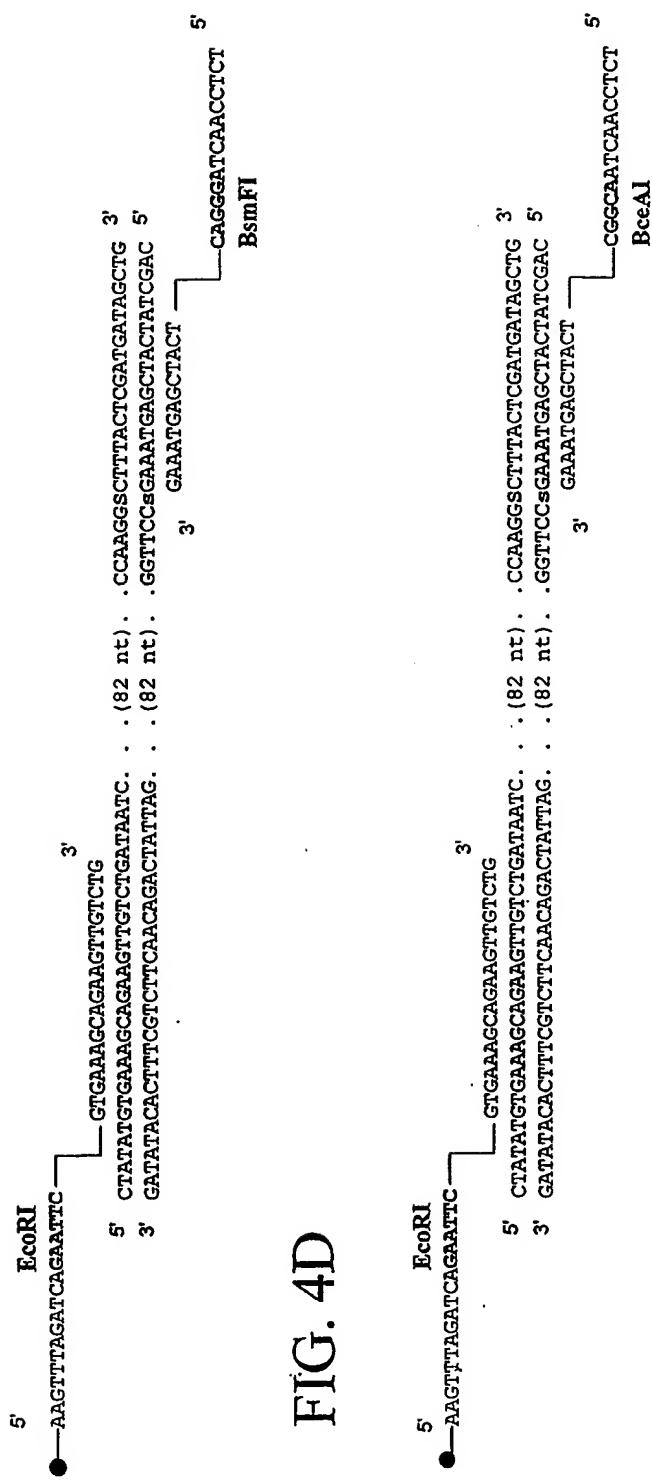


FIG. 4D

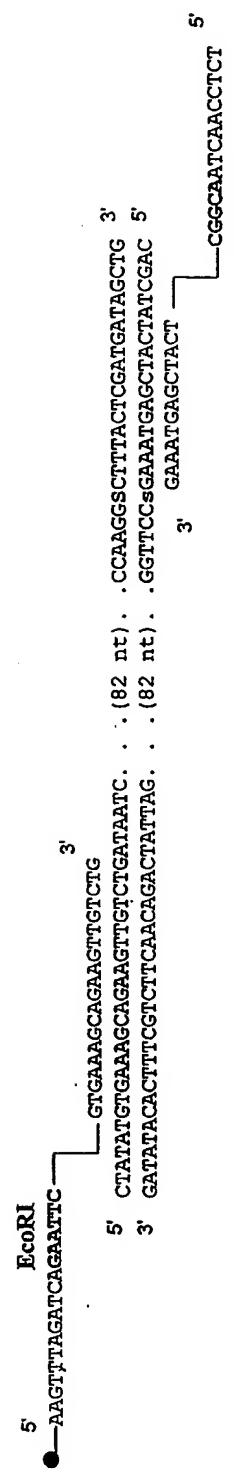


FIG. 5A

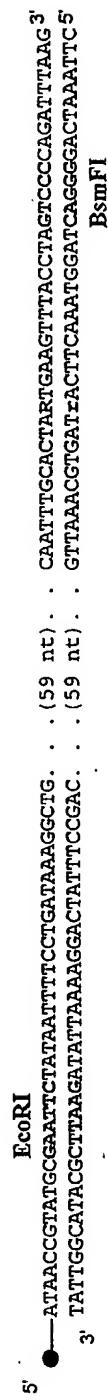


FIG. 5B

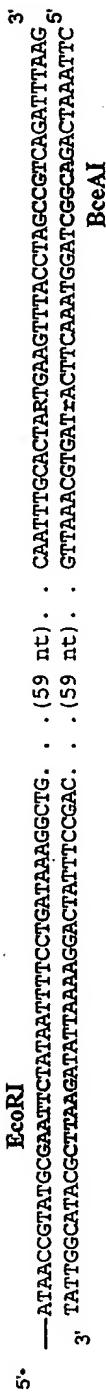


FIG. 5C

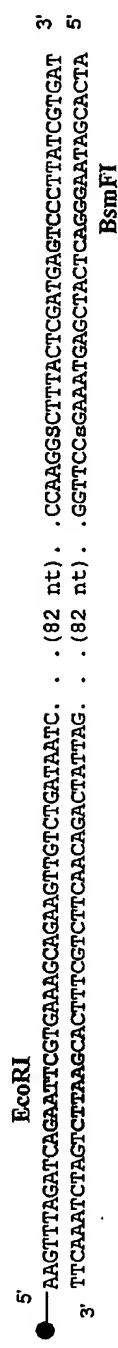


FIG. 5D

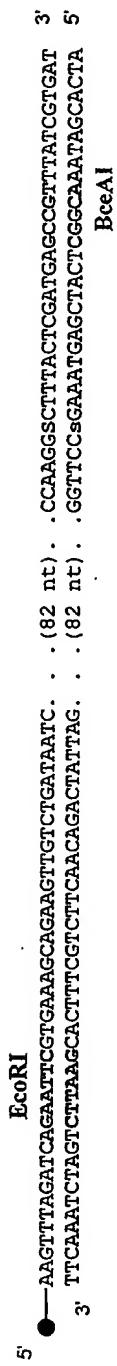


FIG. 6A

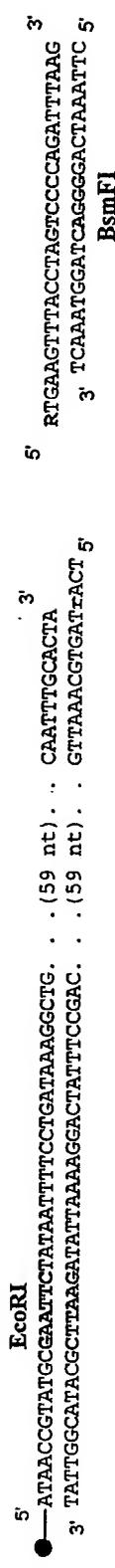


FIG. 6B

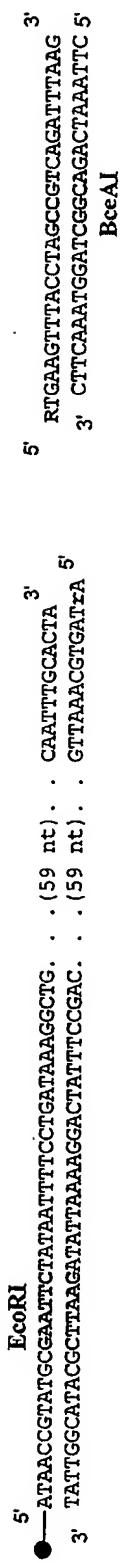


FIG. 6C

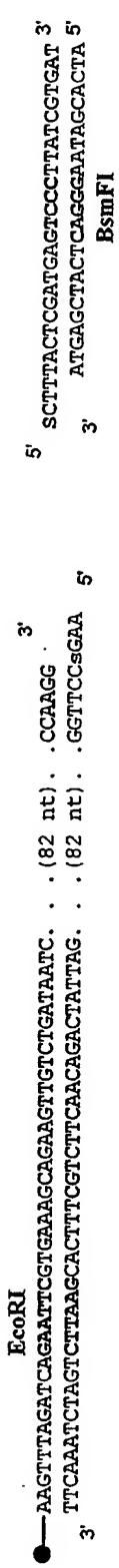


FIG. 6D

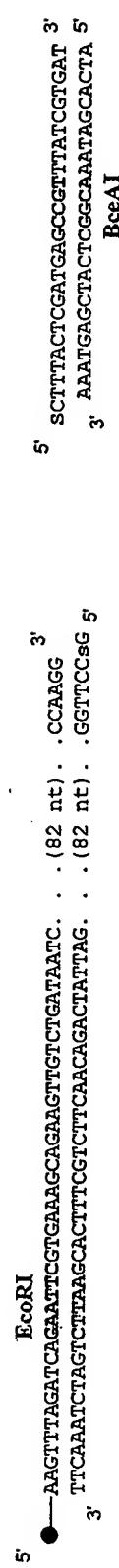


FIG. 7A

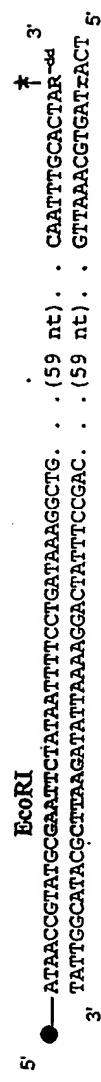


FIG. 7B

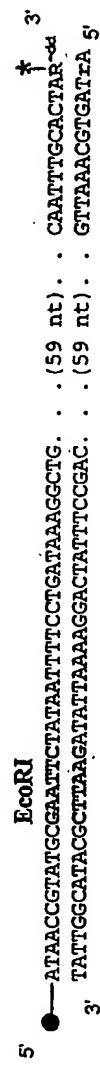


FIG. 7C

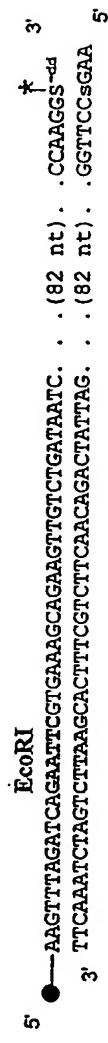


FIG. 7D

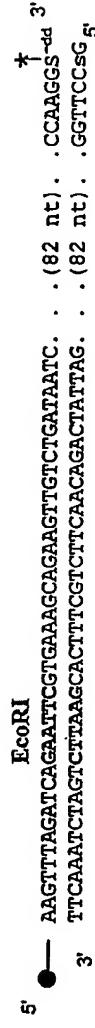


FIG. 7E

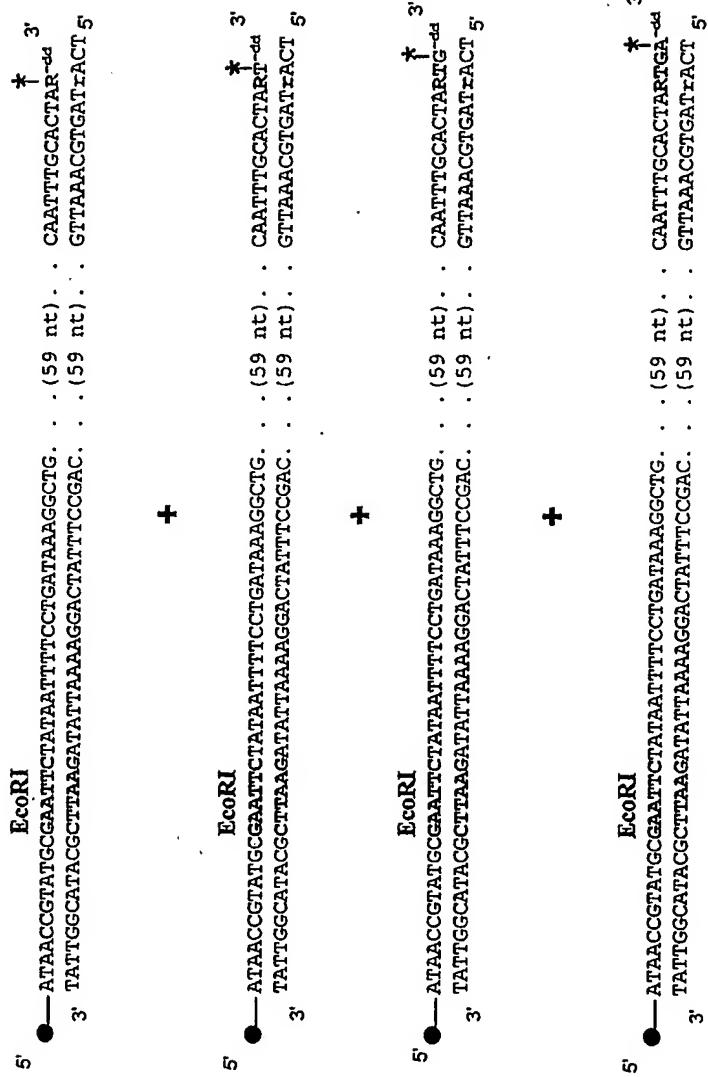


FIG. 8A

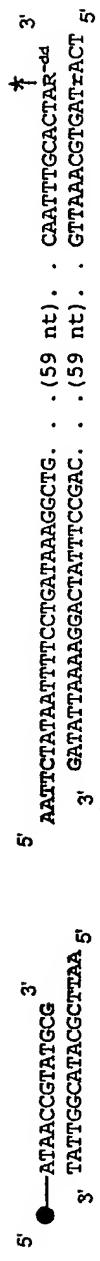


FIG. 8B

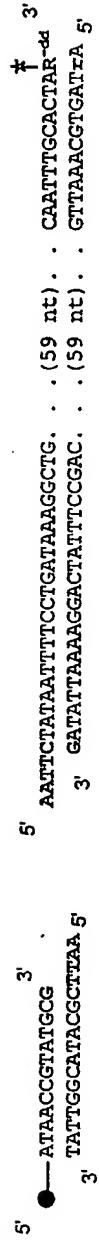


FIG. 8C

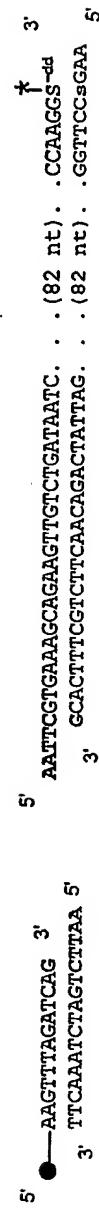


FIG. 8D

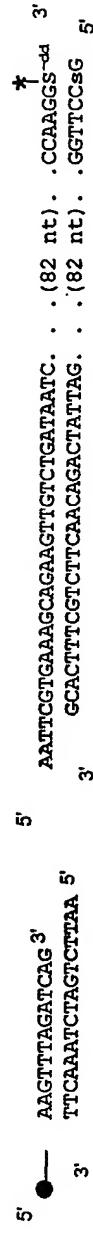


FIG. 9A

FIG. 9B

FIG. 9C

FIG. 9D

| Nucleotides | G A T C | G A T C | G A T C | G A T C | G A T C |
|--|---------------|-------------------|-----------------|---------------|-----------------|
| SNP | HC21S00027 | TSC0095512 | TSC0095512 | TSC0264580 | HC21S00027 |
| Restriction enzyme site on second primer | BceA I | BceA I | BsmF I | BsmF I | BsmF I |
| 5' Overhang | 3' G C s G 5' | 3' G C s G A A 5' | 3' T A r CCC 5' | 3' A r CCC 5' | 3' T A r CCC 5' |
| R | G/A | G/C | G/C | A/C | G/A |
| S | | | | | |

FIG. 10

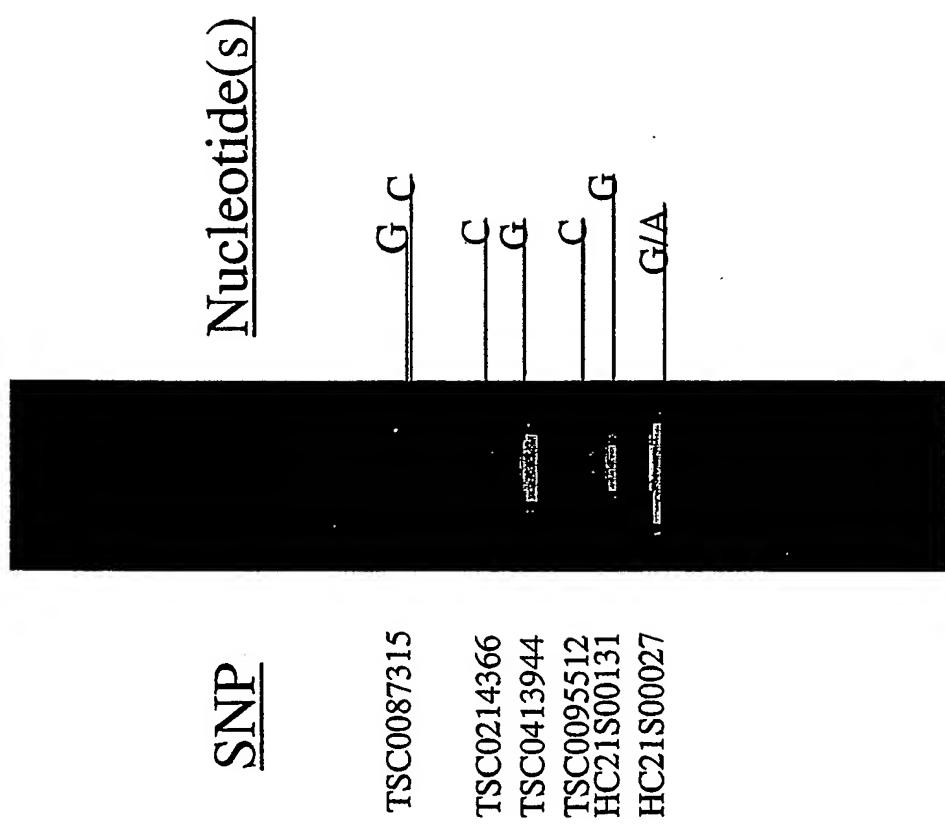
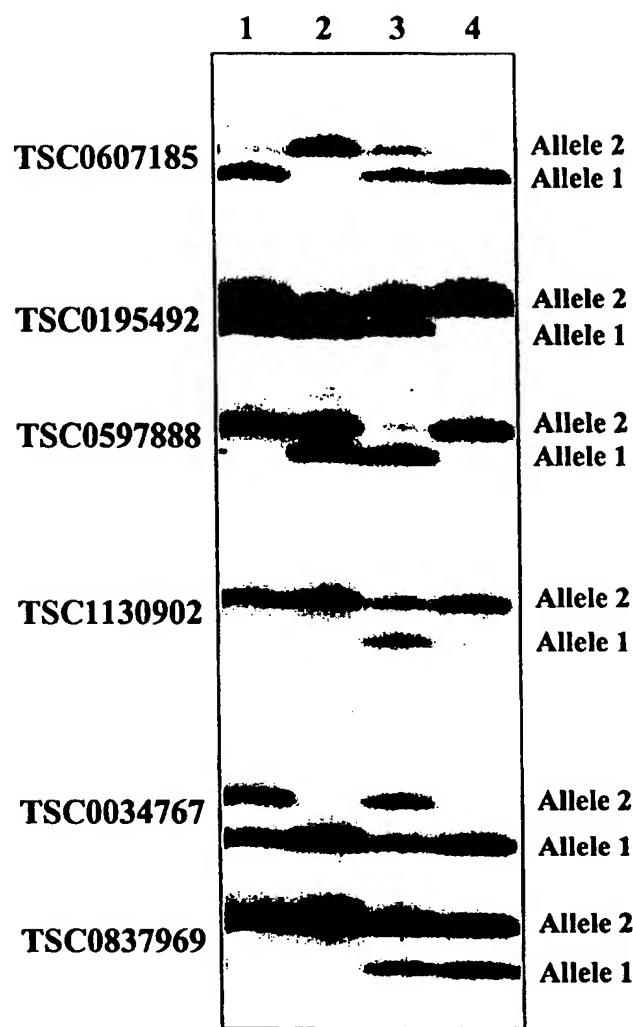


FIG. 11



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/06376

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68;C07H 21/02, 21/04
US CL : 435/6; 536/ 23.1 and 24.3

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 435/6; 536/ 23.1 and 24.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|--|-----------------------|
| A, P | US 2003/0082576 A1 (JONES et al) 01 May 2003 (01.05.2003) see the entire document. | 1-67 |
| A | KWOK, P. Y. Methods for genotyping single nucleotide polymorphisms. Annual Review of Genomics and Human Genetics. 2001, Vol. 2, pages 235-258, see the entire | 1-67 |
| A | SHI, M.M. Enabling large-scale pharmacogenetic studies by high-throughput mutation detection and genotyping technologies. Clinical Chemistry. 2001, Vol. 47, No.2, pages 164-172, see the entire document. | 1-67 |
| A, P | US 6,475,736 B1 (STANTON) 05 November 2002 (05.11.2002) see the entire document. | 1-67 |

Further documents are listed in the continuation of Box C.

See patent family annex.

| | | |
|---|-----|--|
| * Special categories of cited documents: | "T" | later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention |
| "A" document defining the general state of the art which is not considered to be of particular relevance | "X" | document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone |
| "E" earlier application or patent published on or after the international filing date | "Y" | document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | "Z" | document member of the same patent family |
| "O" document referring to an oral disclosure, use, exhibition or other means | | |
| "P" document published prior to the international filing date but later than the priority date claimed | | |

Date of the actual completion of the international search

26 June 2003 (26.06.2003)

Date of mailing of the international search report

17 JUL 2003

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (703)305-3230

Authorized officer

Ethan Whisenant, Ph.D.

Telephone No. (703) 308-0196

PCT/US03/06376

INTERNATIONAL SEARCH REPORT

| C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|---|---|-----------------------|
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| X | US 5,831,065 A (BRENNER) 03 November 1998 (03.11.1998) see the entire document but note especially the figures. | 60 |
| --- | | ----- |
| A | | 1-59 and 61-67 |

INTERNATIONAL SEARCH REPORT

PCT/US03/06376

Continuation of B. FIELDS SEARCHED Item 3:
USPATFULL and EUROPATFULL via EAST; CAplus, Medline via STN

search terms: Genotyp\$ and SNP\$ and/or Single Nucleotide Polymorphism\$ and allele\$; TypeIIs and/ or Type IIs restriction endonuclease\$ and/or enzyme\$, and fill in or fill-in and sequence\$

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.